Summer 8-2002

The Biophysical Basis of Triplet Repeat DNA Expansion in Genetic Disease

Anthony M. Paiva
Seton Hall University

THE BIOPHYSICAL BASIS OF TRIPLET REPEAT DNA EXPANSION IN GENETIC DISEASE

Anthony M. Paiva

This dissertation submitted to the Department of Chemistry and Biochemistry of Seton Hall University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

August, 2002

Mentor: Dr. Richard D. Sheardy

CERTIFICATION

We certify that we have read this thesis and that in our opinion it is adequate in scientific scope and quality as a dissertation for the degree of Doctor of Philosophy.

APPROVED

Mark Chiu, Ph. D., Abbott Laboratories, Inc. (alternate reader)

Stephen Kelty, Ph. D.

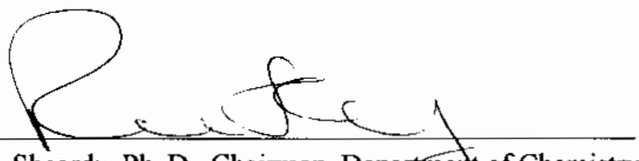
John R. Sowa, Jr., Ph. D.

Richard D. Sheardy, Ph. D.
Chairman, Department of Chemistry

CERTIFICATION

We certify that we have read this thesis and that in our opinion it is adequate in scientific scope and quality as a dissertation for the degree of Doctor of Philosophy.

APPROVED



Richard D. Sheardy, Ph. D., Chairman, Department of Chemistry and Biochemistry

Approved and Accepted for the Department of Chemistry and Biochemistry, Seton Hall University

DEDICATION

To my wife, Laura and kids, Alexa and Nicholas.

ACKNOWLEDGEMENTS

No project of this magnitude could be accomplished in isolation or as the result of an individual effort. To this end, I am indebted those who have played a role in the successful accomplishment of this dissertation.

First I would like to thank former and present members of the Seton Hall University Biochemistry Research Group. Former members, including Michael Hicks and Dave Calderone, for demonstrating the proper use of the CD spectropolarimeter and other laboratory techniques. Stephen Marotta, who developed most of the HPLC gradients and who also wrote the original HPLC DNA purification guide used in this laboratory. Enrique Dillone for valuable discussions about thermodynamics and UV-melting techniques. Current members, such as Jaime Ferreira, for providing computer expertise and helpful hints regarding ITC, Cosimo Antunucci, Carlos Ramirez, and Jack Fergusen for conversations ranging from organic chemistry to thermodynamics. No acknowledgement would be complete without also mentioning the advice of Ms. Diane McGovern-Wolfe, whose insights into the operation of the DNA synthesizer facilitated the synthesis of the CGG 75-mer.

Second I would like to thank my professors, in particular, Dr. McGuinness, Dr. Huchital, Dr. Sheardy, Dr. Chiu, Dr. Kelty and Dr. Sowa for freely providing their time to answer questions and dispense advice. To Dr. Sheardy, I extend a special thanks for all of the useful discussions we had regarding the science of thermodynamics. Such discussions shaped many of the ideas entering this project.

Fianlly, I extend my warmest thanks to my family and dear wife Laura who was a constant source of support and encouragement throughout this project. None of this would have been possible were it not for her dedication and inner strength.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS AND DEFINITIONS

1. A.A; C.C; G.G; T.T; N.N = Nucleotide.Nucleotide on opposite strands, not necessarily base-paired. The common practice is to denote base-paired nucleotides on opposite strands with a center dot as in "A·T". However, the nucleotides referred to with this abbreviation are not necessarily base-paired in a Watson-Crick sense, so a unique abbreviation was adopted to avoid confusion with the abbreviation commonly used in the DNA field.
2. A260 = Absorbance, 260 nm; This abbreviation refers to the assay (as in "..the absorbance was measured with the A260 assay..") and not the actual absorbance at 260 nm (A_{260}).
3. AEC = Averaged Extinction Coefficient
4. AUFS = Absorbance Units Full Scale
5. AXHPLC = Anion Exchange High Performance Liquid Chromatography
6. bp = basepair; This abbreviation is commonly used with a numeral designating a specific length of duplex DNA such as "10-bp" meaning a 10-basepair segment.
7. CD = Circular Dichroism
8. CPG = Controlled Pore Glass
9. CR = Cruciform
10. DCM = Dichloromethane
11. DEPC = Diethylpyrocarbonate; DEPC specifically reacts with the N-7 positions of adenine and guanine in single-stranded DNA molecules.
12. DIBT = Differential Ion Binding Term
13. DMT = Dimethoxytrityl
14. DNA = Deoxyribonucleic Acid
15. DNA_{pol} = DNA polymerase complex consisting of DNA polymerase, DNA helicase, DNA gyrase and associated cofactors.
16. DSC = Differential Scanning Calorimeter/Calorimetry
17. DUP = Duplex
18. EEC = Empirical Extinction Coefficient
19. FPLC = Fast Protein Liquid Chromatography, is the trade-name for a low-pressure chromatographic pumping and detection system marketed by Pharmacia Corp.
20. HP = Hairpin
21. HPLC: High Performance Liquid Chromatography
22. i.d. = inner diameter
23. kD = kilo Dalton
24. KFC = Kinetics Flow Cell
25. KFCC = Kinetics Flow Cell Cap
26. –mer = oligomer; This abbreviation is used to designate the length of an oligonucleotide, such as a 30-mer, or an oligonucleotide 30 nucleotides in length.
27. MWCO = Molecular Weight Cut-Off
28. o.d. = outer diameter
29. ORF = Open Reading Frame
30. PIPES = Piperazine-*N*,*N*'-bis[2-ethane-sulfonic acid]
31. psi = pounds per square inch
32. RCM = Radial Compression Module

33. RPHPLC = Reversed Phase High Performance Liquid Chromatography
34. SEC = Summed Extinction Coefficient
35. SS = Single Strand(ed); meant to imply a denatured DNA state
36. TAE = Tris Acetate EDTA (solution)
37. TBE = Tris Borate EDTA (solution)
38. TCA = Trichloroacetic Acid
39. TEAA = Triethylammonium acetate (solution)
40. TEMED = *N,N,N',N'* -tetrametheylethylenediamine
41. TR = Triplet Repeat or Trinucleotide Repeat
42. TR-DNA = Triplet Repeat Deoxyribonucleic Acid
43. TRS = Triplet Repeat Sequence or Trinucleotide Repeat Sequence
44. UV = Ultra Violet
45. μW = microwatt

LIST OF TABLES

LIST OF FIGURES

1 ABSTRACT

The Biophysical Basis of Triplet-Repeat DNA Expansion in Genetic Diseases

Genetic diseases, such as Fragile-X syndrome, have been linked to the expansion of trinucleotide repeat (a.k.a. triplet repeat) DNA sequences of the 5'-(CNG)n-3' motif during replication. This expansion is related to the formation of self-complementary hairpins that cause reiterative DNA synthesis. While many of the details of the biological mechanism have been defined, the biophysical basis of triplet repeat expansion remains unclear. We examined the structural, thermodynamic and kinetic properties associated with the (CNG)n triplet repeat series where N = A, C, T, or G and n = 4, 5, 10, 15, 25.

On the basis of circular dichroism, native gel electrophoresis and DNA-concentration dependence, that all solitary (CNG)n oligomers formed hairpin structures under physiological conditions. The structural nature of these hairpins was consistent within each oligomer class as a function of length. In contrast, complementary duplexes of these sequences produced CD spectra with classic B-form duplex DNA signatures. Examination of the hairpin to duplex renaturation kinetics revealed a two-fold activation energy difference between short (12-mer) and long (75-mer) oligomers that was consistent with an increase in the transition enthalpy and entropy. In addition, the activation energy for the CAG/CTG series was two-fold higher than that observed for the CCG/CGG series of oligomers. These results imply greater bond breaking as the hairpins approach the transition state that is both sequence and length dependent. On the basis of differential scanning calorimetry and UV-melt techniques, the oligomers displayed a sequence and length-dependent stability profile that reached a plateau at approximately 45 nucleotides or 21 base pairs. As expected, short oligomer sequences were not as stable as long hairpin sequences by approximately 10 kcal/mol. The sequence dependent stability order for the hairpin structures was, from most to least, CGG > CTG > CAG > CCG.

These results are consistent with a model that places the fate of the hairpin formed during replication at the center of the DNA-expansion process. During replication, both short and long hairpin structures form in regions of triplet-repeat DNA (TR-DNA). However, short TR-DNA hairpins have shorter lifetimes due both to the thermodynamic instability of the hairpin and the relatively fast hairpin to duplex renaturation kinetics. The cellular lifetime of the hairpins is such that renaturation to duplex occurs before reassociation of DNA polymerase. In contrast, long TR-DNA hairpins have longer lifetimes due to the relatively higher thermodynamic stability and slower hairpin to duplex renaturation kinetics. The lifetime of these hairpins allows DNA polymerase reassociation prior to duplex renaturation leading to reiterative synthesis. Expansion of TR-DNA sequences is related to both the relative thermodynamic stability and the reassociation rates of the hairpins that form on the daughter DNA strands.

2 INTRODUCTION

2.1 DNA Structure and Biological function

2.1.1 Structural and Conformational Properties of DNA.

The protein-centric view of the biochemical world maintains that, because DNA does participate in cellular biochemistry to the same extent that proteins do, they are uninteresting and unworthy of study. Proteins are, after all, the "center" of the biochemical universe. For example, one can pick at random any number of proteins that interact with DNA as part of their cellular function. For this particular thought experiment, we'll select the glucocorticoid receptor, a member of the nuclear hormone receptor family that binds DNA in ligand-dependent manner. The number of papers that discuss the ligand-binding pocket of the protein, in terms of function or structure far outnumber those papers discussing the role of DNA with respect to function of the protein. For most protein-centrists, the receptor either binds to DNA or it does not and the importance of DNA is somewhere between that of a sodium ion and water – DNA is just another molecule that interacts with the protein, but does little else in terms of function. However, there remains much to be studied about the properties of DNA with respect to cellular function. As the following dissertation will discuss, DNA structure can have enormous effects on protein interaction.

Deoxyribonucleic acid, or DNA, is capable of taking on a variety of structural forms that are dependent upon the DNA sequence and the solution environment. The most common form of DNA found under physiologic conditions within the cell is B-form DNA, which

has two strands composed of a sugar-phosphate backbone that are in an antiparallel orientation. The deoxyribosemonophosphate moiety, which makes up the essential building block of DNA has one of four nitrogenous bases covalently attached to the C-1' or anomeric carbon. These bases, either the pyrimidines cytosine and thymine or the purines guanine and adenine, form the covalent linkage to the anomeric carbon at position N-9 of the purine, or N-1 of the pyrimidine. The four bases can assume two different orientations, *syn* or *anti*, with respect to the deoxyribose sugar. The preferred orientation for the purines, with their extended ring structure, is the *anti* conformation, with the bulk of the ring structure facing away from the deoxyribose ring. A *syn* orientation, with the bulk of the ring structure facing over the deoxyribose is possible under certain conditions, to allow alternative, or non-Watson-Crick base pairing arrangements. For pyrimidines, either orientation is possible due primarily to the relatively small size of the pyrimidine ring. However, the preferred orientation is the *anti* orientation so that the full hydrogen-bonding capabilities of the amines or carbonyl oxygens can be achieved. The nitrogenous bases are capable of forming complementary hydrogen-bond interactions, in a "Watson-Crick" manner, such that guanine forms three hydrogen bonds with cytosine, and adenine forms two hydrogen bonds with thymine. This "Watson-Crick" internal base-pairing arrangement in combination with physiologic salt conditions allows DNA to take on the classic B-form DNA structure, with an antiparallel helix containing a major and minor groove. While the depth of the grooves are approximately the same (8Å) the width of the minor groove for B-form DNA (6Å) is approximately half that of the major groove. The solvent accessibility of the minor groove is therefore lower than that of the major groove.. The internal hydrogen-bonding

arrangement is such that the hydrophobic bases are internal to the molecule, with the highly charged phosphate backbone facing the bulk solution. This conformational arrangement has two effects with respect to DNA structure. First, the hydrophobic interior allows the bases to align and participate in π-π or base stacking interactions between adjacent bases. This stacking interaction contributes to the DNA stability, which is defined by hydrogen bond interactions, base stacking interactions, and counterion condensation (electrostatic interactions). The exposed phosphate groups give rise to a regular and high charge distribution along the DNA backbone, such that there approximately two phosphates every 3 – 4 Å. This high charge density provides the molecular basis for cation condensation that stabilizes the DNA.

2.2 Repetitive DNA and Classes of DNA Repeats

2.2.1 Long Terminal Repeats, Satellites, Microsatellites

Sequences of highly repetitive DNA are scattered throughout most eucaryotic genomes and are often referred to as "satellite-DNA" in reference to the buoyant density of this DNA measured by analytical centrifugation with a cesium chloride gradient.[1] Highly repetitive DNA sequences usually contain asymmetric distributions of A-T and G-C base pairs, thus altering the buoyant density of these regions of DNA. Since the first description of satellite DNA, a variety of sub-types have been identified beginning first with Jeffries who coined the term "microsatellite-DNA".[2,3] In terms of primary structure, satellite-DNA consists of nearly homogeneous segments of DNA containing a common repeat unit (Figure 1).

Figure 1: Example of common repeats.

Two satellite DNA sequences are shown below. Due to the asymmetric distribution of A/T and G/C basepairs, the bouyancy density of these sequences differs from that of genomic DNA.

Crab minisatellite DNA: 5'- ATATATATATATAT-3' (7 repeats, repeat unit AT)

Drosophila satellite: 5'-ATAAATT ATAAATT-3' (2 repeats, repeat unit ATAAATT)

The length of the repeat unit depends upon the source of the DNA, and can vary in size from the 2-base pair (bp) repeat unit observed for crab satellite-DNA, to as high as the 1400-bp repeat observed for bovine satellite-DNA.[4] In humans, a 3-bp repeat is the most commonly observed unit, although longer (5-bp) and shorter (2-bp) sequences have been identified.[5] For example, the telomere repeat unit found within the end of human chromosomes can be from 6 to 12 nucleotides long and can be repeated from 4 to 100 times.[6] In general, the total length of repeats can vary from relatively short (2-1000 bp for microsatellites, 1000-5000 bp for minisatellites) to relatively long (10^5-10^6 bp for common satellites). The number of repeats observed can vary from a few, as observed for the bovine satellite repeat, to several hundred or thousand, as observed for the crab minisatellite repeat.[7] The functional role of repeat sequences are varied, and can be regulatory, as in the case of telomere DNA, structural, or as yet undefined as with triplet repeat sequences.

2.3 Triplet Repeat DNA

2.3.1 Triplet Repeat Sequences

Included within the microsatellite subset is a type of DNA called "trinucleotide repeat sequences" (TRS) or "triplet repeats" (TR) in reference to the three-base pair repeat unit (Figure 2).

Figure 2: The CCG/CGG triplet repeat DNA sequence.

A trinucleotide repeat sequence, CCG, and the complementary strand, CGG, are shown for the Fragile-X repeat. The standard chemical nomenclature of $(CCG)_9$ designation is used in perhaps 80% of the literature regarding triplet repeat sequences. Less commonly used, but still appearing in some journals are the (CCG)9 or the CCG9 designations. In all cases the meaning is the same – a CGG triplet repeat unit that has been repeated 9 times, for a total length of 27 nucleotides.

5'- CCGCCGCCGCCGCCGCCGCCGCCGCCG- 3' (9 repeats, repeat unit CCG)
3'- GGCGGCGGCGGCGGCGGCGGCGGCGGC- 5' (9 repeats, repeat unit CGG)

Designations used: $(CCG)_9$ = (CCG)9 = CCG9

While most trinucleotide repeats are benign, by virtue of localization within untranslated regions of most genomes, several sequences have been implicated in human disease. Of the 64 possible triplet repeat sequences[8] available in the genetic code, seven have been linked to human disease. Four of the seven repeat sequences, CTG, CAG, and CCG/CGG have been linked to myotonic dystrophy (DM), Huntington's disease (HD), and Fragile-X syndrome (FRAX), respectively (Table 1).[9,10,11]

Table 1: The Diseases Caused by Triplet Repeat Sequences.

Genetic Disease/Syndrome	Repeat	Location	Clinical Manifestation
Fragile-X Syndrome, (FRAXA)	CCG	5'-UTR	Mental Retardation
Fragile-X Syndrome, (FRAXE)	CGG	5'-UTR	Mental Retardation
Huntington's Disease	CAG	ORF	Cognitive Impairment
Spinocerebellar Ataxia	CAG	ORF	Ataxia
Spinobulbar Muscular Atrophy	CAG	ORF	Muscle Weakness
Myotonic Dystrophy	CTG	3'-UTR	Muscle Weakness
Freidrich's Ataxia	AAG, CTT	Intron	Limb Ataxia

5'-UTR = 5' untranslated region prior to translated gene
3'-UTR = 3' untranslated region after translated gene
ORF = Open Reading Frame
Intron = intervening sequence of mRNA excised prior to translation

In many instances, these four sequences have also been linked to several diseases other than those described above. In addition, the repeat sequences, AAG and CTT have been linked to cause Freidrich's ataxia (Table 1). From these examples, a common sequence trend emerges - those sequences possessing a CNG motif, where "N" = A, C, G, or T, are particularly prone to cause disease. While other repeat sequences have been found throughout the human genome, the relationship between disease causation and sequence has not yet been established – either there are quiescent triplet repeat syndromes that have yet to be identified, or these sequences do not posses the necessary properties to cause disease.[12] According to Gacy et al., these other commonly observed sequences either do not have the correct threshold energy for structure formation (stability) or are present in lengths too short to cause harm.[12]

2.4 Properties of Triplet Repeat DNA and Genetic Diseases

2.4.1 Properties of Triplet Repeat Sequences and Genetic Diseases

Several generalizations can be made about the properties of triplet repeat sequences and the diseases these sequences cause. The first, and probably most important generalization, is that triplet repeat diseases arise as the result of an expansion in the triplet repeat copy number. Triplet repeat diseases are therefore termed "genetic diseases" because disease causation occurs primarily at the DNA level. The number of triplet repeat codons dictates whether or not a particular individual will be normal, predisposed to disease, or afflicted with one of the previously described conditions. This fact underscores a new mutational mode that differs from what is typically observed for genomic DNA. Most previously described mutations, the so-called "static" mutation, involve changes to the DNA sequence, either through additions, deletions, or substitutions of nucleotides and/or codons. Such mutations involve changes to the sequence or reading frame of the DNA bases (Figure 3).

Figure 3: Static versus dynamic mutations.

A. Static mutations are shown, diagrammatically for a fictitious sequence of DNA. A static mutation involves additions, substitutions, deletions or frameshift mutations that change the sequence of DNA. The changes have been underlined. **B.** Dynamic mutations, shown for the CAG triplet repeat, do not involve changes to the sequence, but rather to the number of repeats. The changes have been underlined.

A. STATIC MUTATIONS:
Wild type sequence: 5'-GGATCCGCTAGT-3'
Addition Mutation: 5'-GGAATCCGCTAGT-3' (reading frameshift)
Deletion Mutation: 5'-GGA_CCGCTAGT-3' (reading frameshift)
Substitution Mutation: 5'-GGAGAAGCTAGT-3' (changes codon)

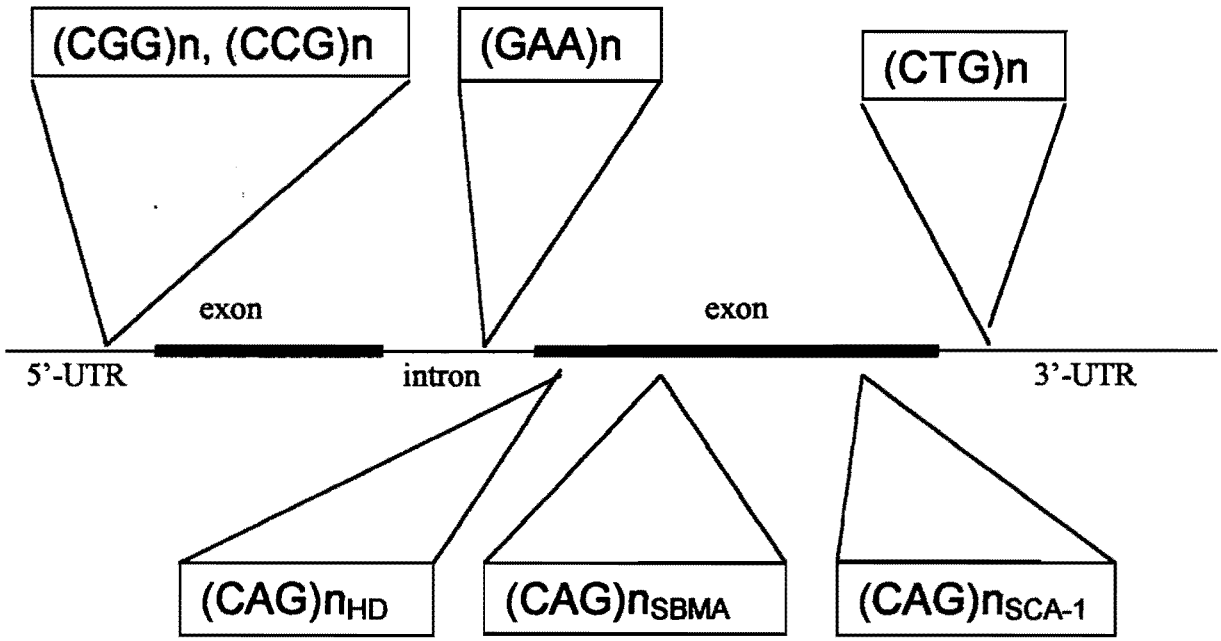
B. DYNAMIC MUTATIONS:
Wild type sequence: 5'-CAGCAGCAGCAG-3'
Dynamically Mutated : 5'-CAGCAGCAGCAGCAGCAG-3'

In contrast, mutations with triplet repeat DNA sequences involve changes in the number of repeat units. For example, a CAG repeat is not changed to a CCG repeat, but rather a segment of DNA containing five CAG repeats will mutate into a segment of DNA containing ten CAG repeats (Figure 3). This novel mutational mode has been called a "dynamic-mutation."

The second generalization is the sequence and the location determines the nature of the disease. Because of this property, the disease mechanism at the DNA level differs for each sequence under consideration. Thus, a common sequence, such as the CAG repeat, can be implicated in several different disease types. If a CAG repeat is located within the open reading frame of the Huntingtin gene[13] the individual is affected with the cognitive impairment typical of Huntington's disease. This same sequence, at the appropriate length, inside of the open reading frame for the androgen receptor will produce spinobulbar muscular atrophy, or if inside of the ataxin-1 open reading frame will produce spinocerebellar ataxia-1 (Figure 4).

Figure 4: The location of triplet repeat DNA sequences.

The relative locations of the triplet repeat sequences are indicated below. The locations shown do imply that all sequences are found in close proximity or within the same gene. The depiction shows the relative locations with respect to intron/exon position, and the relative locations of the triplet repeat sequence within the gene referenced. The CAG insertion for Huntinton's disease, for example, occurs near the 5' terminus of the Huntinin gene sequence, whereas the CAG repeat for SCA-1 occurs near the 3' terminus of the Ataxin gene sequence.



Repeat Sequence	Location	Disease	Normal Length, (repeats)	Premutation Length (repeats)	Affected Length (repeats)
CAG	Huntinin Gene	Huntington's	6-20	—	35
CAG	Androgen Rec.	SBMA	11-25	—	40
CAG	Ataxin-1	SCA-1	6-30	—	40
CCG, CGG	5' UTR	Fragile-X	6-50	60-150	>200
GAA	intron	Freidrich's ataxia	7-20	—	>200
CTG	3'-UTR	Myotonic dystrophy	5-30	50-175	>200

Figure adapted from Reddy, P.S.; Housman, D.E.; **(1997),** *Curr. Opin. Cell Biol.*, 9:364-372.

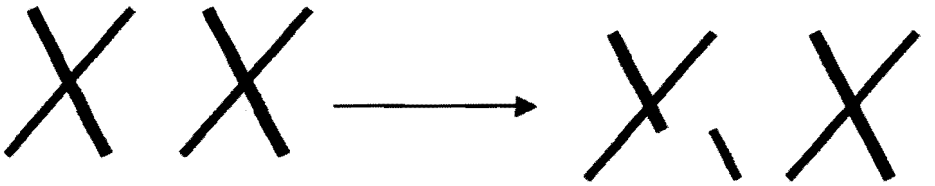
The third generalization is that the disease inheritance is usually autosomal or X-linked. In addition, parental origination effects are common. With the exception of myotonic dystrophy and Fragile-X syndrome, it is common for inheritance of most of the genetic diseases listed to come from the paternal lineage.14 With respect to the CCG series of triplet repeats, the extrahelical cytosine-cytosine mismatch15 renders the CCG-region of DNA susceptible to hypermethylation by DNA mismatch repair enzymes. Hypermethylation of the C-rich regions targets the DNA for excision, and in many instances renders the X-chromosome susceptible to single-strand cleavage by DNA repair enzymes. This single-strand cleavage of the chromosome weakens the DNA backbone, and renders the strand susceptible to strand scission. In the case of a female child, the second X-chromosome compensates for the loss of genetic information. However, in the case of a male child, the lack of a compensatory "arm" on the Y-chromosome leads to disease causation (Figure 5).

The fourth generalization about triplet repeat sequence disease is that the severity of the outcome and the time of onset is dependent upon the number of repeat units an individual carries.[14] For example, with respect to Fragile-X syndrome, the length of the CCG or CGG repeat is between 20-50 repeat units in normal individuals whereas, in disease-affected individuals, this number can be in excess of 250 repeat units.[12,16,] With respect to the CAG triplet repeat, the disease threshold is much lower. Whereas normal individuals typically possess between 5 to 10 repeats,

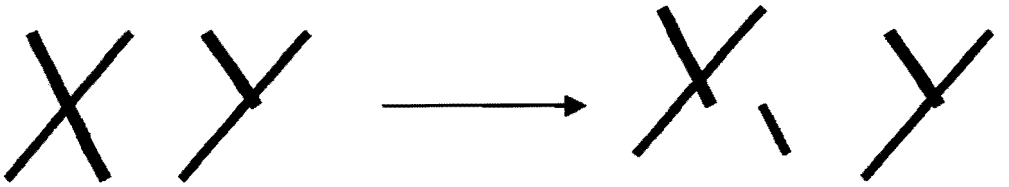
Figure 5: A schematic view of X-linked inheritance.

A fragile-X sequence located on one arm of the X-chromosome can lead to strand scission and loss of genetic information. **A.** The loss of genetic information, located on one arm of the X-chromosome is compensated by the sister X-chromosome. **B.** The loss of genetic information is not compensated by the sister Y-chromosome.

A.



B.



diseased individuals possess between 25 or more repeats.17 This phenomenon is related to the fifth and sixth generalizations to follow.

The fifth generalization is that the repeat number becomes unstable once a critical number of repeat units have been attained. Low copy numbers have been associated with stable transmission from parent to child over many generations.[14] However, under certain conditions that remain undefined, the repeat number may expand because of errors in normal DNA replication. Once the repeat number expands to a critical threshold, the likelihood of further expansion increases. This phenomena, is related to the sixth generalization, that the likelihood or predisposition for further expansion in the repeat number is dependent upon the number of repeats a carrier possess. The higher the number of repeats as a child, the more probable the repeat number will be expanded in a nonlinear fashion to a higher value later in life or from parent to child.

Seventh, approximately seven out the 64 possible triplet repeat sequences are known to cause disease. These known sequences are CNG (where N = A, C, T, or G), AAG, and CTT. Confounding this observation is the fact that other triplet repeat sequences have been found within the human genome but none to date have been linked to disease. This raises the obvious question about why these, over all other sequences can result in disease.

The eighth property is that the expansion frequency for all triplet repeat sequences is not equivalent among all sequences. It has been shown that CCG sequences are repaired

with lower efficiency than the CTG or CGG sequences.[18] The observed expansion frequency is CAG>CTG>CCG>CGG, and has been thought to derive from the inability of DNA repair enzyme to recognize and repair the lesion in the chromatin structure as well as stability effects arising within the DNA repair site.

Finally, expansions in repeat number arise during DNA replication, but disease arises from the downstream effects of the DNA insertions. There are many theories about the expansion mechanism, including the slip-DNA model and the superhelicity model (described below), but none have been definitively proven. What is currently known about the expansion mechanism is that hairpin formation is thought to occur during replication on the daughter strand DNA.

In general, triplet repeat diseases can be classified into two different groups, designated Type I and Type II diseases.[19] Type I diseases, such as Huntington's disease, Spinocerebellar ataxia, and Spinobulbar Muscular Atrophy generally involve the central nervous system and are related to expansions of CAG triplet repeats within the open reading frames of encoded proteins. This polyglutamine region within the protein causes aggregation within the cell due to the high charge density between the amino acids of adjacent proteins. It is thought that aggregation leads to neuronal cell death due to formation of plaques and/or aggregates within the cell. Type II diseases[19] involve multiple organ systems, and are related to expansions of CCG, CGG, and CTG triplet repeats outside of the protein-coding region. With respect to the CCG, CGG and CTG repeats, the localization and mechanism of disease causation differs from that observed

for the CAG series of repeats. Verkek and others have mapped the location of most CCG/CGG repeats to the 5'-untranslated region of the FMR-1 gene (Figure 4).[20,21] Because of the location of CGG sequences within the 5' untranslated (but not untranscribed) regions of expressed genes, it is easy to believe that CGG triplet repeat sequences should have little or no effect on gene transcription and downstream translation. However, the current understanding proposed by Siomi and coworkers is that 5' region of the FMR1 gene encodes an RNA-binding protein.[22] Insertions of a CGG triplet repeat segment decreases expression of the FMR1 gene by decreasing RNApol binding. It is also thought that self-complementarity of the triplet repeat sequences can lead to the formation of alternative mRNA structures, such as hairpins or quadruplexes, that in turn affect the ability of mRNA processing enzymes to cleave the introns from the mRNA transcript prior to gene assembly.[16]

From this broad view of the properties of both triplet repeat DNA sequences, and triplet repeat diseases emerges a few trends worth noting. First, only a handful of triplet repeats, out of the many possible combinations of triplet repeats found, are capable of causing genetic disease. The number of repeats can remain at a low and stable copy number for years until an unknown event triggers an expansion in the number of repeats. Eventually the number of repeats reaches a critical number beyond which further expansion is very probable. As a whole, these data suggest a possible link between the DNA structure formed by triplet repeat sequences during replication, in the case of expansion mechanisms. Because structure is closely related to the thermodynamics, or the stability of the structure formed, there is the possibility for an additional relationship between

disease and sequence, or disease and structure. For this and many other reasons a number of research groups have attempted to address the biophysical basis of triplet repeat expansion diseases.

2.5 The Biophysical Properties of TR-DNA

2.5.1 Overview

The questions that researchers within the triplet repeat DNA field have attempted to address to date include identification of the structures formed by these sequences both *in vitro* and, wherever possible, *in vivo* (Zheng,[15] Petruska,[28] Gacy[29] and others). Related to the structural studies were investigations about the thermodynamic properties of each triplet repeat sequence (Petruska,[28] Gacy,[29] and others). Other research groups have attempted to address the properties of long versus short triplet repeat sequences in order to gain an understanding as to why short segments of triplet repeat DNA remain stable for generations (Gacy[12] and others). Additional investigations attempted to address the underlying mechanism, in terms of the cellular events during replication, that could lead to expansion (Gellibolian[30] and others). Finally, the hairpin to duplex association kinetics of complementary triplet repeat DNA stands were studied to determine the relationship between structure and the rate of reassociation. What follows therefore is a brief review of the literature regarding triplet repeat DNA.

2.5.2 Structural Studies: Mitas, Zheng, Petruska

Zheng and coworkers studied the NMR solution structure of (CNG)4 oligonucleotides, where N is A, C, T or G in 10 mM sodium phosphate, 100 mM sodium chloride, 0.1 mM

EDTA at pH 6.8 – 8.3.[15] Their results indicated that all CNG triplet repeats form antiparallel duplexes that can readily undergo structural transitions due to increased flexibility within the sequence. Each sequence, however, differed in the preferred alignment of the mismatched bases. In particular, CCG repeats formed a novel DNA structure, called e-DNA, that is a staggered duplex with an antiparallel alignment. In this structure, two CG residues were base-paired in a normal Watson-Crick sense, followed by an unpaired C residue (schematically shown as 5'-CCGCCG-3'). To accommodate this unpaired base, the DNA helix adopted an elongated structure (hence e-DNA), where the pyrimidine base is located across the minor groove facing the bulk solution. For CGG repeats, the two CG residues were followed by an unpaired G residue in the anti orientation on one strand and the syn orientation on the opposing strand. This reorientation allowed the mismatched guanines of the two strands to form alternative hydrogen bonding arrangements that conferred stability to the hairpin oligomer. Gacy et al. also studied these sequences by NMR and found essentially the same results.[12] Previous work by Mitas utilized chemical probes to determine the solvent-accessibility of the mismatched TT base pairs of the CTG triplet repeat.[23] They found that the T-T mismatches in CTG repeats are stacked within the helix, protecting these bases from $KMnO_4$ modification. In addition, they also found that A.A mismatches in CAG are not well stacked (protected) based upon the ease of diethylpyrocarbonate (DEPC) modification. While these observations were consistent with the observations made by Gacy and Zheng, the work by Mitas differs in two respects. First Mitas shows that CGG repeats can form tetraplex hairpins, and second, that CCG can adopt two distinct hairpin conformations, one of which is the e-DNA conformer described by Zheng. The

observation that CGG repeats could form stable tetraplexes, however, was first described by Fry and Loeb[24] and underscores some of the variation observed in DNA structure. Later work by Fry, however, indicated that the tetraplex structure observed for CGG repeats is possibly a foldback hairpin.[25] Chen et al. also studied the solution structure of CCG and CGG repeats by NOESY NMR techniques and found essentially the same results.[26] However, their gel electrophoresis data indicated that the duplex to hairpin transition is dependent upon the length, the DNA concentration and the salt concentration. For a given length of DNA, the duplex was favored at low salt, whereas the hairpin was favored at high salt concentration. As the DNA-length increased the hairpin form became the favored species even at low salt concentrations. Their work, however, does not study all of the disease relevant triplet repeats and is therefore applicable only to the limited sample set studied.

Nadel and coworkers studied the behavior of the CGG triplet repeat sequence at low (10 mM sodium chloride) and physiologic salt (100 mM sodium chloride) concentrations, with DNA concentrations between 0.1 and 1 μM.[27] The gel electrophoresis results indicate that the sequences formed unimolecular fold-back hairpins, with a 4 or 7-nucleotide loop. Depending upon the sequence and length, they also proposed the possibility for internal "bulge" formation within the stem of the hairpin structure.

To summarize, triplet repeat DNA sequences can spontaneously form unique and stable structures, due primarily to the self-complementary nature of the DNA and in part to the inherent DNA-strand flexibility that allows alternative hydrogen-bonding arrangements.

The type of structure formed depends upon the solution conditions that include parameters such as the pH, the salt concentration and the DNA concentration. Studies where the DNA concentration was high, as is the case with NMR experiments, found evidence of duplex structures, or mismatched duplexes. Studies utilizing lower DNA concentrations such as gel electrophoresis, found evidence of hairpin formation.

2.5.3 Thermodynamic Studies (Petruska, Gacy, Santalucia)

The thermodynamic properties of various triplet repeat DNA sequences were measured by several research groups to date. However, because the sequences and solution conditions were appropriate to the specific questions asked by the researchers, many cross-study comparisons of thermodynamic properties has proven difficult. Thus, the complete data set of thermodynamic properties for all seven of the known disease-causing triplet repeat sequences remains undefined.

One of the first to study these sequences from a thermodynamic perspective was Petruska, et al. who found that the order of stability as measured by the free energy, ΔG, and enthalpy of the 30-mer stability in 20 mM or 120 mM sodium chloride was CTG>GAC = CAG> GTC (Table 2, Table 3).[28] In addition, the melting temperature (T_m) of the CAG and CTG triplet repeats changed very little as the sequence length changed from 30bp to 90bp (Table 4). Although the enthalpy was expected to show a 3-fold increase from a 30bp to 90bp sequence, the observed change was only 1.4-fold. They concluded, on the basis of both the T_m and ΔH, that the hairpins formed under these conditions are not "typical" hairpins, where stem length correlated as a function of DNA length. Rather, these hairpins contained multiple mismatched basepairs that can account

Table 2: The ΔG of the hairpin/duplex to coil transition.

A comparison of the free energy for the hairpin/duplex to coil transition from each reference cited. Units for ΔG are kcal/mol, and have been corrected so that the sign of the free energy is for the single-strand to hairpin/duplex direction.

Sequence	Author	ΔG
CAG4	Zheng	-6.3
CCG4	Zheng	-8.5
CTG4	Zheng	-4.3
CTG8	Zheng	-11.3
CAG10	Gacy	-2.2
CAG10	Petruska	-1.1
CTG10	Gacy	-4.3
CTG10	Petruska	-2.2
CAG25	Gacy	-5.3
CCG25	Gacy	-3.2
CGG25	Gacy	-12.4
CTG25	Gacy	-7.8
CAG30	Petruska	-1.4
CTG30	Petruska	-3.1

Petruska, J.; Arnheim, N.; Goodman, M.F.; **(1996)**, *Nuc. Acids. Res.*, 24:1992-1998.
Zheng, M.; Huang, X.; Smith, K.; Yang, X.; Gao, X.; **(1996)**, *J. Mol Biol.*, 264:323-326.
Gacy, A.M.; McMurray, C.T.; **(1998)**, *Biochem.*, 37:9426-9434.

Table 3: A comparison of the ΔH of the HP/DUP to SS transition.
A comparison of ΔH/basepair for the hairpin to coil transition from each reference cited. Because each author studied a triplet repeat of differing length, the reported values were normalized by dividing the reported ΔH by the number of basepairs. Units for ΔH are kcal/mol.

Sequence	Author	ΔH
CAG10	Gacy	-28.9
CAG10	Petruska	-38
CTG10	Gacy	-65.6
CTG10	Petruska	-55
CAG25	Gacy	-60.8
CCG25	Gacy	-88.1
CGG25	Gacy	-120
CTG25	Gacy	-100
CAG30	Petruska	-50
CTG30	Petruska	-76

Petruska, J.; Arnheim, N.; Goodman, M.F.; **(1996)**, *Nuc. Acids. Res.*, 24:1992-1998.
Zheng, M.; Huang, X.; Smith, K.; Yang, X.; Gao, X.; **(1996)**, *J. Mol Biol.*, 264:323-326.
Gacy, A.M.; McMurray, C.T.; **(1998)**, *Biochem.*, 37:9426-9434.

Table 4: A comparison of the melting temperatures of the oligomers.

The number in parenthesis refers to the length of the DNA sequence. For all measurements the salt concentration was between 1-10 mM.

	T_m	T_m	T_m	T_m
Sequence	**Zheng (12mer)**	**Mitas (45mer)**	**Gacy (75mer)**	**Petruska (90mer)**
(CAG)n	23	38	50	46
(CCG)n	49	30	49	--
(CGG)n	--	75	75	--
(CTG)n	5	47	51	51

Petruska, J.; Arnheim, N.; Goodman, M.F.; **(1996)**, *Nuc. Acids. Res.*, 24:1992-1998.
Zheng, M.; Huang, X.; Smith, K.; Yang, X.; Gao, X.; **(1996)**, *J. Mol Biol.*, 264:323-326.
Gacy, A.M.; McMurray, C.T.; **(1998)**, *Biochem.*, 37:9426-9434.

for the apparent decrease in ΔH. Zheng, on the other hand, found a different trend in the stability of $(CNG)_4$ repeats.[15] The solution conditions were 10 mM sodium phosphate, 100 mM sodium chloride, 0.1 mM EDTA, with the pH varied from 6.8 to 8.3, and the DNA concentration at 10 μM. The stability order, based upon free energy calculations derived from T_m measurements (van't Hoff enthalpy) was found to be CGG>CAG>CTG>CCG. In addition, they also found, based upon NMR data, that the flexibility of the mismatched basepairs in CCG and CAG repeats would not allow stable and paired conformations with each other. While this observation provides NMR support for the conclusion drawn by Petruska, there remains some difference in the stability order of the triplet repeat sequences studied, due perhaps to experimental conditions. In a more recent work, a partial study of triplet repeat thermodynamics was conducted by Gacy and McMurray.[29] Like their predecessors, they find little difference in stability, as indicated by the change in enthalpy for the HP to SS transition for {(CTG)25 and (CTG)10} or {(CAG)25 and (CAG)10}. Furthermore, the reported free energies for the HP to SS transition differ from those derived by Petruska and Zheng (Table 2). They conclude that the hairpin structures contain many mismatched basepairs. They also found that hairpin formation is favored with longer sequences, and duplex formation is favored with shorter sequences.

In summary, many studies with respect to the thermodynamics of triplet repeat sequences have been conducted, with conflicting results. One general trend detected is that the stability of the hairpin, or other structure formed by these sequences, does not correlate with sequence length. Furthermore, no single study is comprehensive enough, or similar

enough in solution conditions, to allow comparisons across the entire set triplet repeat sequences.

2.5.4 Kinetic Analysis, (Gacy)

The hairpin (HP) to duplex (DUP) reannealing kinetics of triplet repeat sequences had received little attention until a relatively recent report by Gacy and McMurray.[29] Their work focuses in part upon the process through which triplet repeat hairpins reassociate with the parental DNA strand during replication. Prior to this publication, it was known that hairpin formation within self-complementary triplet repeat DNA sequences occurred during DNA replication. It was theorized (but never proven) that eventually the hairpins would reanneal to the parental duplex, not only because the duplex form is the most stable form of DNA, but because transmission of dynamic mutations requires transmission in a cellularly stable form. Any hairpins that remain after replication appear as a lesion on the chromosome and can possibly be excised by the DNA repair mechanisms. The rate at which the hairpins reanneal to the parental DNA, forming complementary duplex DNA, can determine the subsequent disease state.

For the kinetics portion of this work, they studied the duplex formation rates for CAG10 with CTG10, CAG25 with CTG25, and CCG10 with CGG10 in 10 mM Piperazine-*N*,*N*'-bis[2-ethane-sulfonic acid] (PIPES), 100 mM sodium chloride, 0.1 mM EDTA, pH7.0. They found that the kinetic properties defined the difference between long and short repeat sequences. Both sets of sequences formed hairpins with similar rates of formation (data not shown) that was independent of length. However, the long sequences

reassociated to duplex with slower kinetics than did the short sequences. The observed rate constants, measured for the hairpin to duplex reaction was 8500/mol*sec for the (CAG)25/(CTG)25 pair, and 40,000/mol*sec for the (CAG)10/(CTG)10 pair. This represents an increase of over 5 times the rate observed for the 75-bp sequence. They conclude, based upon these results, that long triplet repeat sequences have sufficient cellular lifetimes, due to the slow reassociation kinetics to allow reiterative-type DNA synthesis. The hairpin remains prior to DNA polymerase (DNA_{pol}) association, allowing the polymerase complex to synthesize a previously replicated region of DNA.

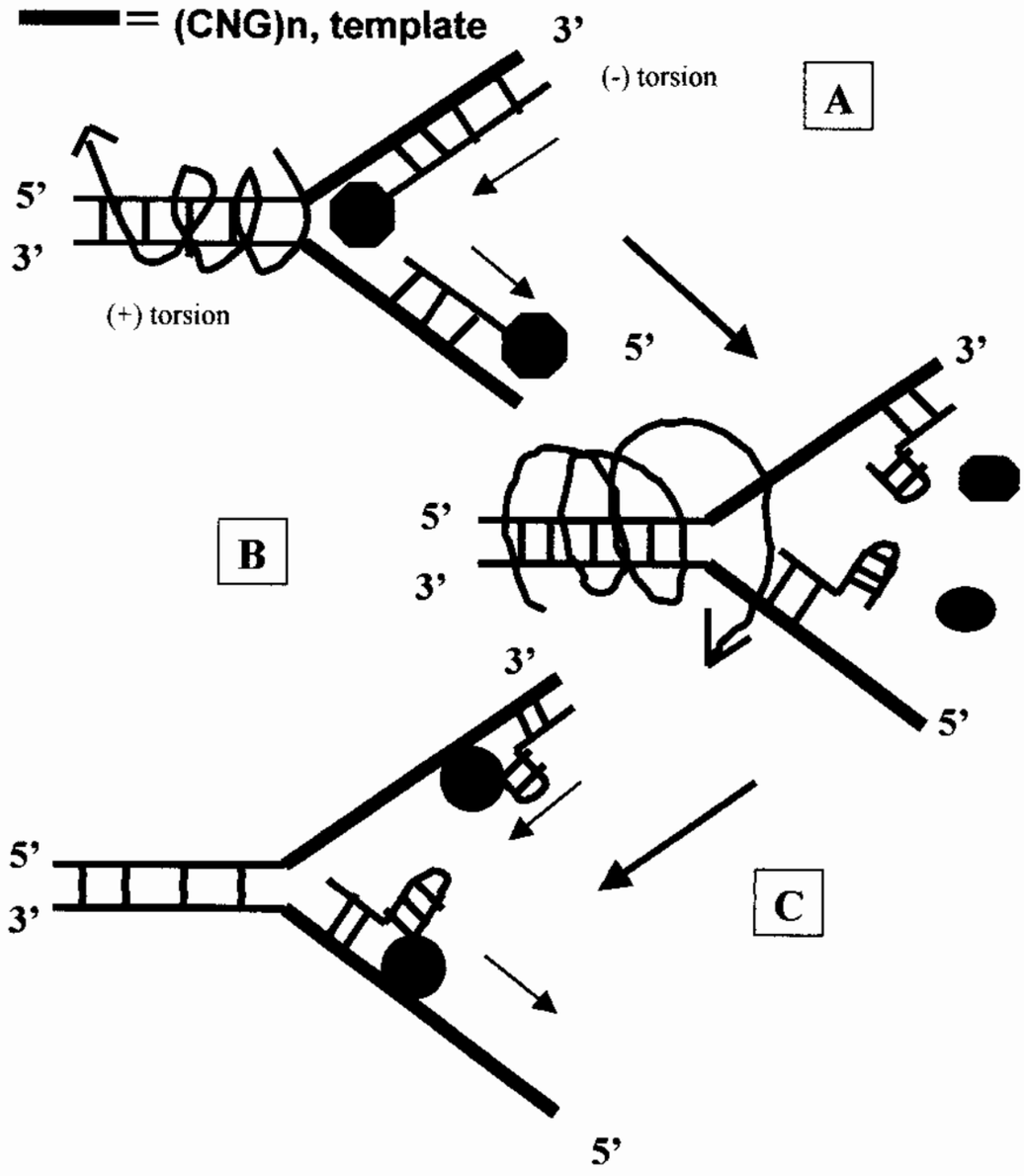
While their work provides insights into the reaction kinetics, and a possible explanation why long repeats are more prone to expansion, they do not define all of the kinetic parameters necessary for a reaction mechanism. In addition, they do not compare the kinetics of all triplet repeat classes across all lengths. The kinetics of (CTG)25/(CAG)25 was compared to (CTG)10/(CAG)10, but they did not compare the kinetics of (CCG)25/(CGG)25 with that of (CCG)10/(CGG)10 - the kinetics of (CCG)25/(CGG)25 was compared against (CTG)25/(CAG)25. Therefore, the relationship between the hairpin-to-duplex association kinetics and triplet repeat length remains to be fully defined. In addition, the observation that triplet repeat sequences with high repeat copy numbers reanneal to duplex with slower kinetics than sequences with low copy numbers raises the intriguing possibility of a length-dependent difference in the energy of activation for these sequences. If there is a fundamental difference in the activation energies then there will also be the possibility of different reannealing pathways both in terms of sequence context and length.

2.5.5 Mechanistic Studies (Bacolla, Gellibolian, Wells)

Currently, the debate now centers around what structures form *in vivo*, and, most importantly, what effects these structures have on the replication process. To this end, several theories have been put forth. For example, Bacolla and coworkers show that the inherent flexibility of triplet-repeat DNA can lead to an accumulation of superhelicity (commonly known as writhe).[30,31] Taking this property into consideration, Gellibolian proposed that unwinding of the parental double-stranded DNA near triplet repeat regions introduces a high degree of positive supercoiling ahead of the DNA_{pol} complex. [1,32] Eventually the accumulated superhelical strain becomes so great that the DNA_{pol} can no longer proceed with unwinding of the double-stranded DNA.[33] At this point, the polymerase must either stop synthesis but remain associated to the DNA, or dissociate from the twisted DNA strand. The superhelical strain can be partially relieved by dissociation of the DNA_{pol} complex from the DNA and, according to their theory, is the only pathway that allows the opportunity for formation of secondary structures within the daughter strands of triplet repeat regions of DNA. Eventually the DNA polymerase will reassociate with the parental DNA strand. If hairpin formation occurs prior to DNA_{pol} association, then the DNA_{pol} will attach to a location upstream (3' with respect to the parental strand position) of the dissociation point, leading to reiterative synthesis and expansion (Figure 6). The slip-DNA causes reiterative synthesis and expansion because the DNA_{pol} complex does not move forward.

Figure 6: The reiterative synthesis model proposed by Gellibolian.

A. The DNA_{pol} complex begins to unwind the DNA duplex containing a triplet-repeat sequence. Because of the flexibility inherent in these sequences, the DNA accumulates positive superhelicity ahead of the complex and negative superhelicity behind the complex. **B.** Eventually, the accumulated writhe stalls the DNA_{pol} complex, leading either to incomplete replication (stalled DNA synthesis) or dissociation of the DNA_{pol} complex. Dissociation of the DNA_{pol} complex allows release of the superhelical torsion, and formation of hairpin (or other) structures prior to association of the DNA_{pol} complex. **C.** The DNA_{pol} complex reassociates downstream (with respect to the template strand) and begins DNA synthesis reiteratively.



The mechanism of Gellibolian, if correct, suggests that the properties of the triplet repeat hairpins, once formed, are critical to the nature of the expansion mechanism. If the hairpin formation occurs, then either the structures are stable or they are not. Stable hairpin formation will eventually lead to reiterative synthesis, whereas unstable daughter-strand hairpins will eventually reassociate with the parental strand leading to no expansion in repeat number.

2.6 Project Aims, Purpose, and Rationale.

2.6.1 Overview and Purpose

The following items summarize the relevant biophysical data collected thus far: The expansion of triplet repeat DNA sequences occurs during replication. The expansion of these sequences occurs because of the formation of structure, possibly hairpins on the daughter strands of the lagging and leading strand synthesis. The structures form as the result of the energy imparted by the superhelical writhe, or simply from the extrusion of unwound DNA during the unwinding process. Sequences with copy numbers at or above the threshold are more prone to expansion of repeat number than sequences with low copy numbers. All of this data, together with the mechanism proposed by Gellibolian and the kinetics data provided by Gacy suggest that the mechanism of triplet repeat expansion may center on the nature of the hairpin structures that form. The fate of the hairpins - whether or not they remain as metastable entities or reanneal to the parental DNA strand - determines in large part the fate of the expansion mechanism. If the hairpins reanneal prior to reassociation of the DNA polymerase complex, then no expansion occurs. If however, the DNA hairpins remain, then the DNA polymerase will

reassociate upstream, and proceed with a reiterative DNA synthesis. The properties of the DNA hairpins, or other structures that form as a result of replication are therefore critical to the expansion mechanism (Figure 7). To date, no research group has fully defined the kinetic, structural and thermodynamic properties of all triplet repeat sequences. We sought to define these properties and determine what relationship might exist between long versus short strands, as well as the sequence context of the CNG series of triplet repeats from a biophysical perspective (Figure 8). In short, what are the structures formed with respect to length and sequences context? Can the observed differences in expansion behavior between high and low copy number be attributed to differences in thermodynamic stability? Is there a difference in the activation energy for the hairpin to duplex reassociation process with respect to length or sequence?

2.6.2 The Sequences and Lengths Under Study

Sequences of the (CNG)n class of triplet repeats, where N = A, C, G, or T and n = 4, 5, 10, 15, or 25 were studied for this project. These sequences were selected so that sequence context effects (CAG versus CCG, etc.) could readily be compared across different types of triplet repeat sequences. For this reason, and due to the scale of such a project, the AAG and CTT sequences were not studied. The lengths were selected for many reasons. First, 25 repeat units represents a length that is within the disease causation threshold for the CAG series of triplet repeats, and just under the disease causation threshold for the CCG, CGG, and CTG triplet repeats. Thus, direct comparisons could be made between the stabilities, structures, etc., with respect to the CAG series of triplet repeats in both a disease-causing and nondisease-causing context.

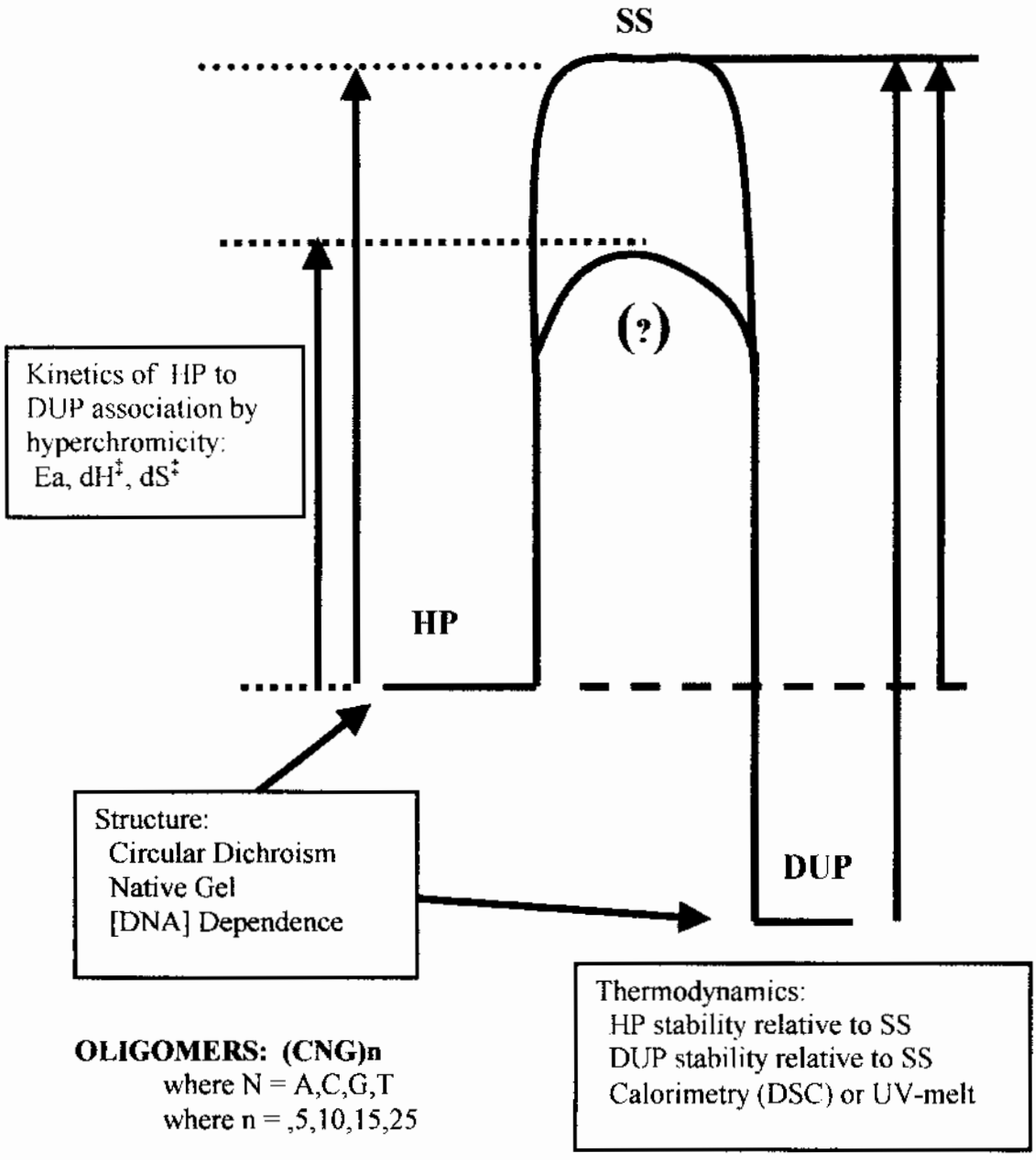
While the 25-repeat length for the other three triplet repeat sequences (CCG, CGG, CTG) is below known disease causation thresholds it was believed that the general trends with respect to the structures, stabilities, etc., between the short and long sequences would also hold true for lengths at the disease causation threshold (50 repeats is the average threshold). The intermediate lengths of 5, 10 and 15 repeat units, were selected to ensure that sufficient date points were collected within the experimental measurement range so that useful trend information could be obtained. The shortest length (4 repeat units) was selected initially as control sequence, because this sequence was not expected to contain any significant secondary structure, but was later found to form a stable hairpin. The 4 and 5 repeat unit oligomers were designed to answer specific questions about the structure and/or stability of the hairpin loop.

2.6.3 Characterization of the Structural Properties

To study the structures formed by these sequences, all oligomers were studied either as solitary oligomers (in the absence of the complementary Watson-Crick strand)[34] or as complementary duplexes through several orthogonal techniques – circular dichroism, native gel electrophoresis, and concentration dependence. The secondary structure of DNA, also defined by the number of base pairs per turn, the rise per base pair, the handedness of the helix and the distance of each base from the helix axis, is polymorphic and dependent upon the solution environment. Circular dichroism spectroscopy is sensitive to these changes and can therefore be utilized to monitor the changes in DNA secondary structure.[35] Changes to any of the base-base parameters, such as base-to-base distance, or stacking interactions, brought about by changes in either the solution conditions or changes in the DNA [illegible]

Figure 8: The biophysical characterization of the expansion mechanism.

The biophysical properties of the (CNG)n oligonucleotides, where N = A, C, T or G, and n = 4, 10, 15, 25, were measured from a variety of perspectives with the techniques described. The oligonucleotides were used as models for the *in vivo* behavior of the triplet repeat DNA sequences. The sequences and lengths chosen allow for direct comparisons across the thermodynamic, kinetic and structural perspectives.



While the 25-repeat length for the other three triplet repeat sequences (CCG, CGG, CTG) is below known disease causation thresholds it was believed that the general trends with respect to the structures, stabilities, etc., between the short and long sequences would also hold true for lengths at the disease causation threshold (50 repeats is the average threshold). The intermediate lengths of 5, 10 and 15 repeat units, were selected to ensure that sufficient date points were collected within the experimental measurement range so that useful trend information could be obtained. The shortest length (4 repeat units) was selected initially as control sequence, because this sequence was not expected to contain any significant secondary structure, but was later found to form a stable hairpin. The 4 and 5 repeat unit oligomers were designed to answer specific questions about the structure and/or stability of the hairpin loop.

2.6.3 Characterization of the Structural Properties

To study the structures formed by these sequences, all oligomers were studied either as solitary oligomers (in the absence of the complementary Watson-Crick strand)[34] or as complementary duplexes through several orthogonal techniques – circular dichroism, native gel electrophoresis, and concentration dependence. The secondary structure of DNA, also defined by the number of base pairs per turn, the rise per base pair, the handedness of the helix and the distance of each base from the helix axis, is polymorphic and dependent upon the solution environment. Circular dichroism spectroscopy is sensitive to these changes and can therefore be utilized to monitor the changes in DNA secondary structure.[35] Changes to any of the base-base parameters, such as base-to-base distance, or stacking interactions, brought on by changes in either the solution conditions or changes in the DNA itself should produce a corresponding change to the CD spectrum.

In addition, the CD spectra of each repeat sequence as function of repeat length was measured over a range of salt concentrations and, in a separate study, temperatures. The CD spectra from these experiments can be overlaid and compared to not only determine if a transition exists but also to determine the structural end states of the DNA as a function of salt concentration and temperature. Such information could prove invaluable for determination of the beginning structural form of the DNA for subsequent kinetics experiments. However, the information gained from CD measurements can define the structure formed only if the sample spectrum is reasonably comparable to a known reference spectrum. For this reason, the oligomers listed need to be studied by other techniques.

With respect to the solitary oligomers, the structure-to-coil transition should be unimolecular if the strands form solitary hairpins, and bimolecular if the strands form mismatched duplexes, and tetramolecular if the strands form quadruplexes. The molecularity of the association process can be determined with the theoretical methods described by Marky and Breslauer[36] with the following equation (Equation 1)

Equation 1: $$\mathbf{1/T_m = \{(n-1)R/\Delta H\}\ln C_t + [\Delta S-(n-1)R\ln 2+R\ln(n)]/\Delta H}$$

where n is the molecularity, C_t is the total oligomer concentration, and R is the universal gas constant. A plot of $1/T_m$ vs. $\ln(C_t)$ yields a plot with slope equal to $(n-1)R/\Delta H$ and an intercept of $\Delta S/\Delta H$. For a unimolecular transition (hairpin) the slope of such a plot will be equal to 0, and for all other associations (e.g. bi, trimolecular) the slope will be equal to some negative value. Thus, through measurement of the concentration dependence of

the melting temperature, the molecularity of the transition, and hence the structure could be readily determined.

To further clarify the nature of the structures formed, native gel electrophoresis was also be utilized. Native gel electrophoresis separates on the basis of both size and charge, and can give information about the relative size of biomolecules with good resolution. With proper data interpretation, native gel elctrophoresis can give a basis for comparison between different sizes of the various species involved. When separated under native conditions, the solution structures of the DNA oligomers can be readily compared to clarify if the beginning state is a hairpin, mismatched duplex or tetraplex structure.

2.6.4 Characterization of the Thermodynamic Properties

The thermodynamics of DNA transitions can be monitored with one of two techniques - optical absorbance as a function of temperature, (also called UV-melt analysis) or calorimetry. In principle, UV-melt analysis relies upon the fact that ordered DNA, such as that found in hairpins, duplex etc, has a lower extinction coefficient than denatured (random coil) DNA. In structured DNA, π-π stacking interactions arise as each base is stacked on top of the preceding base of the helix, leading to a hyperchromic effect - the UV extinction coefficient of single stranded (random coil) DNA is higher than the extinction coefficient of native (structured) DNA.[37] Hyperchromicity can therefore be utilized as a tool to monitor the progress of the structure to coil transition as a function of temperature. From these melting data the van't Hoff transition enthalpy can be calculated through application of the theoretical framework first described by Marky and Breslauer.[36] In addition, the model-dependent van't Hoff enthalpy can also be used to

determine the nature of the structure-coil transition through comparison with the calorimetrically derived transition enthalpy. The van't Hoff transition enthalpy will match the magnitude of the calorimetrically determined enthalpy for a true two-state interaction (all native or all random coil) and be smaller if a two-state interaction does not occur. The theoretical basis described by Marky and Breslauer is valid only if the sequences under study undergo a two-state interaction. To confirm this fact, heat capacity for all of the triplet repeat sequences was measured by differential scanning calorimetry. Such measurements provide a model-free approach to obtaining thermodynamic information and therefore represent the most accurate method. Differential scanning calorimetry can also be modeled with appropriate algorithms to produce a van't Hoff transition enthalpy that can be used for direct comparison of the transition enthalpies derived by UV-melt techniques. Therefore, an internal control between the two different methods of data collection can readily be achieved.

Ultimately, the UV-melt derived van't Hoff thermodynamic data and the calorimetric data will be used to correlate the stability of the triplet repeat as a function of length and sequence context. In addition the stability, as defined by the free energy (ΔG) of the structure-coil transition, for each triplet repeat sequence under study will be examined to determine what parameters govern hairpin stability. Is the observed stability or instability related to an increase in enthalpy, entropy or both?

To date, the relationship between salt type, [salt] and triplet repeat structure/stability remains undefined in spite of the fact that binding of ions to DNA can dramatically affect

the conformation, stability and thermodynamics of a DNA system. Sodium ion can shield, through counter ion condensation, the negatively charged phosphates along the DNA backbone. This reduces the phosphate-phosphate repulsion, and allows the DNA to assume a more compact, as well as more stable structure. Therefore, another goal of this project is to investigate the relationship between sodium ion concentration and DNA structure, as well as the effect [salt] has on DNA stability. As suggested in the structural characterization section, measurement of the thermodynamic behavior of these sequences will not only provide information with respect to salt effects, but will also define the behavior (beginning and end state) of these oligomers for kinetics experiments. Sequences that release more sodium ion upon denaturation, can assume more compact structures due to the increased shielding of the negatively charged phosphate backbone. This in turn may contribute to difference in stability that can be related to the observed disease phenomena.

2.6.5 Characterization of the Kinetic Properties

As described earlier, (section 2.5.5) hairpins form on the lagging and leading daughter strands during DNA replication as a result of daughter DNA slippage, accumulation of superhelicity, etc. Eventually the hairpins will reanneal to the parental duplex. Measurement of the kinetics of this process can provide information about the energies of activation and possibly the mechanism for this process if any of the sequences undergo a fundamentally different transition state. In addition, the data can be replotted as an Eyring plot, to determine the transition enthalpy and entropy. Such information can provide insights about the nature of the intermediates formed during this process. Additional mechanistic information can also be obtained by studying the effects of

sodium concentration on the reassociation kinetics of complementary hairpins based upon the work of Azivonas and Kearns.[38] Comparison of the reassociation kinetics across all triplet repeat sequences, with respect to length and context, will yield the underlying reasons for the observations described by Gacy and McMurray.[29] This information, in addition to the thermodynamic and structural information will explain some of the reasons for the observed expansion behavior across the different triplet repeat DNA sequences.

3 MATERIALS AND METHODS

Author's Note:

The procedures in this chapter have been described in detail primarily to provide a working guide for future generations of graduate students. Each procedural section has been roughly subdivided into two to three sections – "Essential Principles," " Sample Prep and Procedures," and where applicable, "Data Analysis." The Essential Principles section attempts to answer the following questions – First, what is the basis of this technique? Second, how is the technique applicable to DNA? Finally, why is this technique used and what are some of the strengths and weaknesses of the technique? The discussions within this section were not meant to be comprehensive, but rather as brief introduction to the principles used so that future graduate students can get started as quickly as possible. For more comprehensive treatment of the subject, the reader is referred to the source literature referenced in the many endnotes included in the section. The "Sample Preparation and Procedures" section describes in detail the procedures used to both prepare and analyze the samples. Included in this section are many endnotes that

either describe technical hints or clarify the underlying reasons for the procedures adopted. In many instances, these endnotes are the result of questions asked by beginning graduate students and undergraduates alike. In many sections the data analysis and interpretation of data was made part of the procedural section because much of the data analysis could not be easily separated from the procedural details. For example, when conducting an RPHPLC purification it is necessary to analyze and interpret the chromatographic data prior to proceeding to the next step. In some sections, such as modeling of DSC data, the data analysis procedures could be separated or needed to be treated separately and are presented as a section unto itself.

3.1 Definition of Equipment and Essential Principles of Operation.

3.1.1 High Performance Liquid Chromatography (HPLC) System

Most analytical and preparative separations were performed with a Waters HPLC system. Occasionally separations were performed with alternate HPLC systems that utilized different hardware or data collection methods. In spite of these differences all systems utilized identical gradients, buffers, and HPLC columns. Therefore, for the purposes of this project, all systems were considered identical. Two Waters model 510 pumps, configured for high-pressure mixing, were controlled through a Waters model 680 automated gradient controller. Absorbance was detected using a Waters Lamda-Max model 481 variable wavelength detector tuned to 280 nm for preparative HPLC or 260 nm for analytical HPLC.[39] Absorbance versus time data were collected with a Nelson Analytical series 900 A/D converter, and all data processed with TurboChrom version 4.1 software. For analytical purposes, the injection loop size was either 50 µl or 100 µl

whereas for preparative purposes the injection loop size was 2000 µl. The RPHPLC column, a Waters µ-BondaPak C-18, 8 mm by 100 mm Radial-Pak column was housed inside of a Waters model RCM 810 radial compression module (RCM). The RCM was pressurized according to the manufacturer's instructions. For analytical AXHPLC, a Hychrom Inc., SynchroPak AX300, 4.6 mm by 250 mm anion exchange column was utilized. All HPLC solvents were HPLC-specification and were thoroughly degassed under vacuum and/or sonication prior to use. All buffer solutions were prepared with either reagent-grade salts or HPLC-grade solvents. All buffer solutions were filtered through a 0.45 µm filter, and were vacuum degassed just prior to use.[40]

3.1.2 Circular Dichroism Spectropolarimeter (CD)

All CD measurements were made with an AVIV Associates CD spectropolarimeter model 62-ADS equipped with a multiple cell turret, a Peltier heating/cooling device and nitrogen purging capabilities. The heat-sink for the Peltier thermoelectric heating/cooling device was an IsoTemp water bath model 9100 (Fisher Scientific) maintained at 20 °C.[41] The nitrogen source was a liquid nitrogen Dewar container (NJ Welding Supply) configured for gas siphoning. The spectropolarimeter was interfaced to a DOS-based computer running AVIV version 3.14 software for command/control, data collection and data processing functionalities.[42] Long-range UV quartz cuvettes with a width of 2 mm and path length of 10 mm (Helma Scientific) were used for all studies. All cuvettes were scanned, with water, from 320 nm to 210 nm prior to use to detect any strain in the quartz surfaces. Cuvettes with noise levels above +/- 5 millidegrees were not used for CD spectrophotometry.

3.1.3 Differential Scanning Calorimeter (DSC)

All heat capacity measurements were made with a Nano-II power-compensation differential scanning calorimeter from Calorimetric Sciences Corporation. The calorimeter was equipped with dual 299 µl platinum-alloy capillary cells mounted inside of a Peltier heating/cooling block.[43] The calorimeter was interfaced directly to a Windows-based computer with "DSCRun" version 2.1.1 software (Calorimetric Sciences Corporation) for both data collection and DSC control. The data was blank corrected, integrated, and modeled with "CpCalc" version 2.1 (Calorimetric Sciences Corporation) and Origin version 6.1 software. All samples were thoroughly degassed using a vacuum dessicator (without dessicant) and filtered prior to measurement. Every month the background residual scan was recalibrated using water. Every six months, the cells were balanced.

3.1.4 Glass Vacuum Degasser

A glass dessicator (Fisher Scientific) without dessicant[44] was connected to a vacuum source capable of achieving a vacuum level of 25 inches of mercury. The dessicator contained a purge valve near the top through which the vacuum source connected.

3.1.5 PolyPakII Cartridge Column (PPII)

The PolyPakII cartridge column (Glen Research Inc.) contained a proprietary reversed-phase stationary phase bonded to a polymeric resin solid support. Typically, the column was connected to a 10 ml syringe and all column eluents were forced through the column by hand-pressure. Large-volume wash steps were accomplished using a laboratory-made vacuum manifold consisting of a side-arm vacuum flask equipped with a Luer fitting on

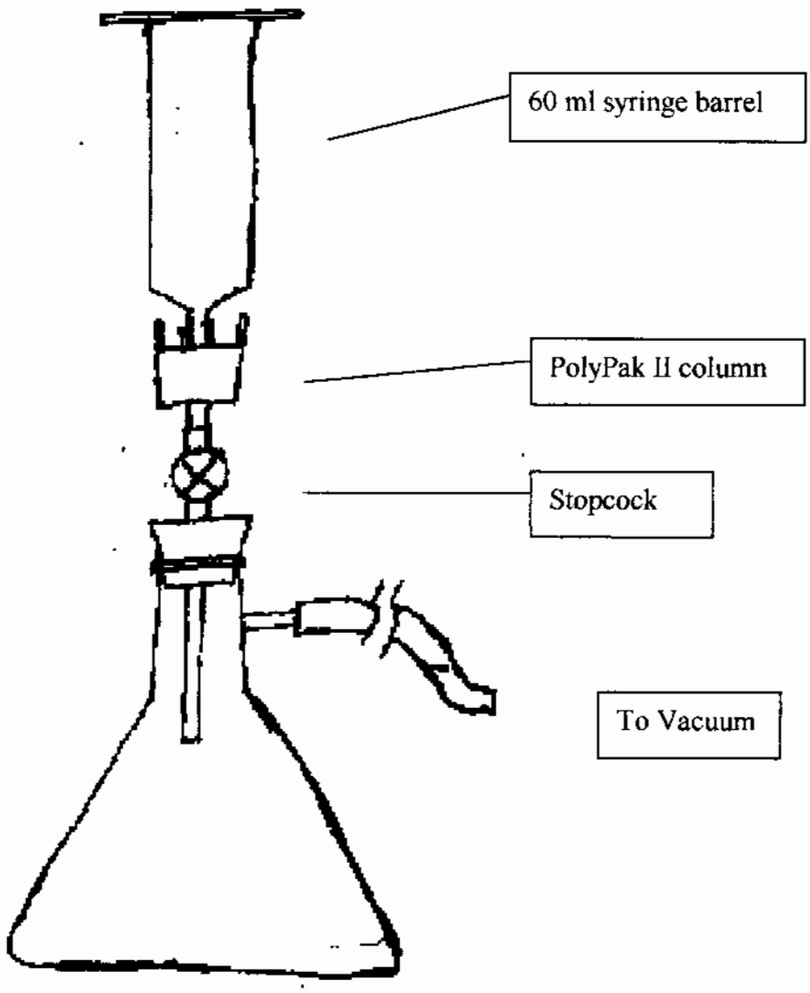
top (Figure 9). The vacuum flask was attached to either a vacuum pump or water aspirator pump.

3.1.6 DNA Synthesizer

An Applied Biosystems Inc. (currently ABI-Perkin Elmer) automated DNA synthesizer model 380B was used for all DNA synthesis reactions. Downstream of the ABI synthesizer was a trityl monitor composed of a Pharmacia "Fast Protein Liquid Chromatography" (FPLC) Single-Path UV-1 absorbance detector (filter, 280 nm) connected to a Pharmacia FPLC Single-Path UV-1 Absorbance Detector Controller set for 2.0 absorbance units full scale (AUFS). The analog output of the detector/controller pair was connected to a Pharmacia strip chart recorder set for 0.5 mm/min. The synthesizer utilized β-cyanoethylphosphoramidites and associated chemistry either from Glen Research, Inc. or CruaChem, Inc.[45] Anhydrous acetonitrile[46] was obtained from Aldrich Chemical Company, Inc. and reagent grade ammonium hydroxide from Fisher Scientific. The low-moisture acetonitrile was anaerobically transferred, with the aid of a Schlenk manifold, from the Aldrich Sure-Seal bottles to previously dried and argon-purged amber glass bottles fitted with rubber septa.[47] The acetonitrile was stored under dessication and dated as to the date of transfer. Any acetonitrile older than 6 months post-transfer was recycled for general laboratory purposes. The reagents of the ABI model 380B synthesizer were delivered to the column and maintained on the synthesizer under a constant supply of low-moisture, high-purity argon (NJ Welding Supply) at 70 psi.[48] Safety-coated 10 X 130 mm screw-capped culture tubes (Corning Scientific, cat no. 69825-13) were used for collection of the crude oligonucleotide solution, and all deprotection reactions.

Figure 9: The PolyPak II rinsing/washing vacuum manifold.

The manifold is constructed with a side-arm vacuum flask, a rubber stopper, and a Luer fitting. The Luer fitting, which is a standard syringe connection, enables the use of a 60 ml syringes as the buffer reservoir. Flow rates, approximately 1 - 2 ml/min are controlled by a miniature stop-cock valve located below the Luer fitting. (Note: There are commercially available stopcocks with a built-in Luer fitting. These can be obtained from many suppliers of low-pressure chromatography fittings such as Pharmacia, AllTech, and Upchurch Scientific).



3.1.7 Econo-Pump System

A modified Econo-Pump system from Bio-Rad, Inc. enabled a semi-automated approach to oligomer purification with PPII cartridge columns. A buffer "reservoir" (60 ml syringe barrel), eluent switching valve (Upchurch Scientific), and peristaltic pump (Bio-Rad model EP-1), were connected in series ahead of the PPII or other chromatography column. An ultra-violet absorbance detector (Bio-Rad model EM-1) set for 1.0 absorbance unit full scale at 254 nm, was connected downstream of the PPII column (Figure 10, next page). The analog output of the absorbance detector was connected to a Kipp-Zoenen model BD-11 strip chart recorder set for 2.0 mm/min. The flow rate of the pump, controlled by both the internal diameter of the peristaltic tubing and by the rotational speed of the pump rotor, was set to approximately 1 to 2 ml/min. This flow rate was slightly faster than the manufacturer's recommended flow rate of 1 drop/sec. When pumping manually, the PPII column was connected to a 30 ml syringe in place of the Econo-Pump (Figure 11). All downstream connections and settings remained the same as specified for the semi-automated approach.

3.1.8 Gel Electrophoresis Equipment

Two different electrophoresis systems were used, depending on the number of samples to be analyzed or the type of gel analysis. Small numbers of samples were analyzed with a modified ExCell II electrophoresis unit from Novex, Inc. The electrophoresis chamber lid was bored to accept ¼ inch internal diameter thin-walled Tygon tubing in both the front and rear buffer compartments (Figure 12). Modification of the

Figure 10: The PolyPakII/Econo Pump low-pressure chromatography system.

A schematic drawing of the PolyPak II cartridge column connected in series with an Econo-Pump System for the ammonium hydroxide wash, and water washing steps. The main advantage of the Econo Pump system for DNA purification with the PolyPak II columns is the real-time monitoring of purification. The flow rate of the pumping system is controlled in part by the internal diameter of the peristaltic pump tubing, and the rotation rate of the pump head. The flow rate is between 1 -2 ml/min. (Note: Not drawn to scale)

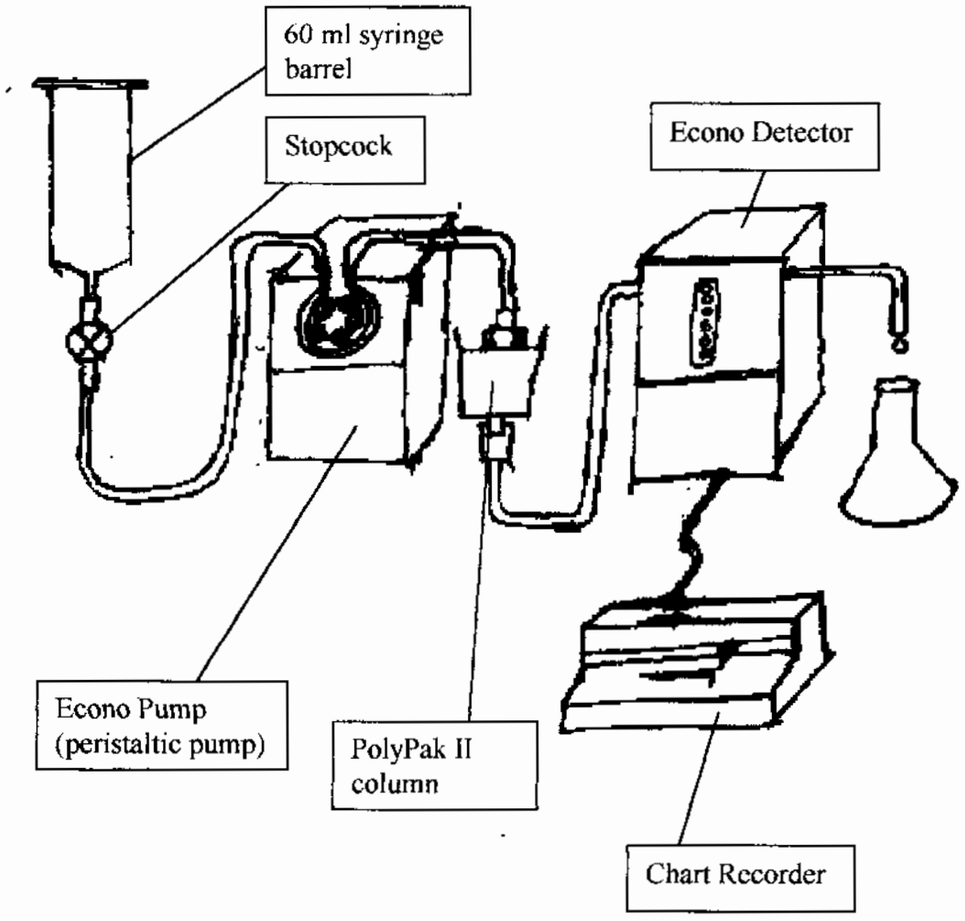


Figure 11: The PolyPak II/Econo system configured for manual pumping.

A schematic drawing of the PolyPak II cartridge column connected in series with an Econo-Pump Detector for the detritylation reaction. It is possible to eliminate the Econo-Pump, by manually pushing the solutions through with a suitable syringe. This configuration is suitable for detritylation reactions, where the flow rate must be faster than 1 – 2 ml/min and within a certain length of time. (Note: Not drawn to scale)

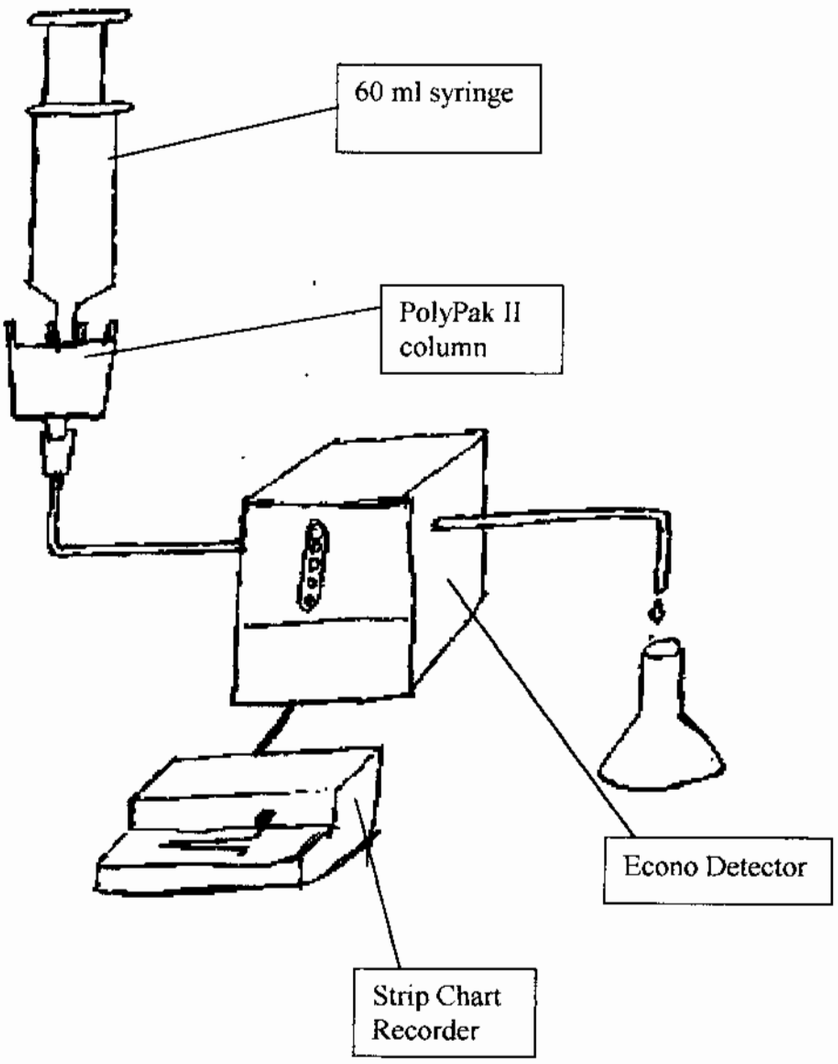
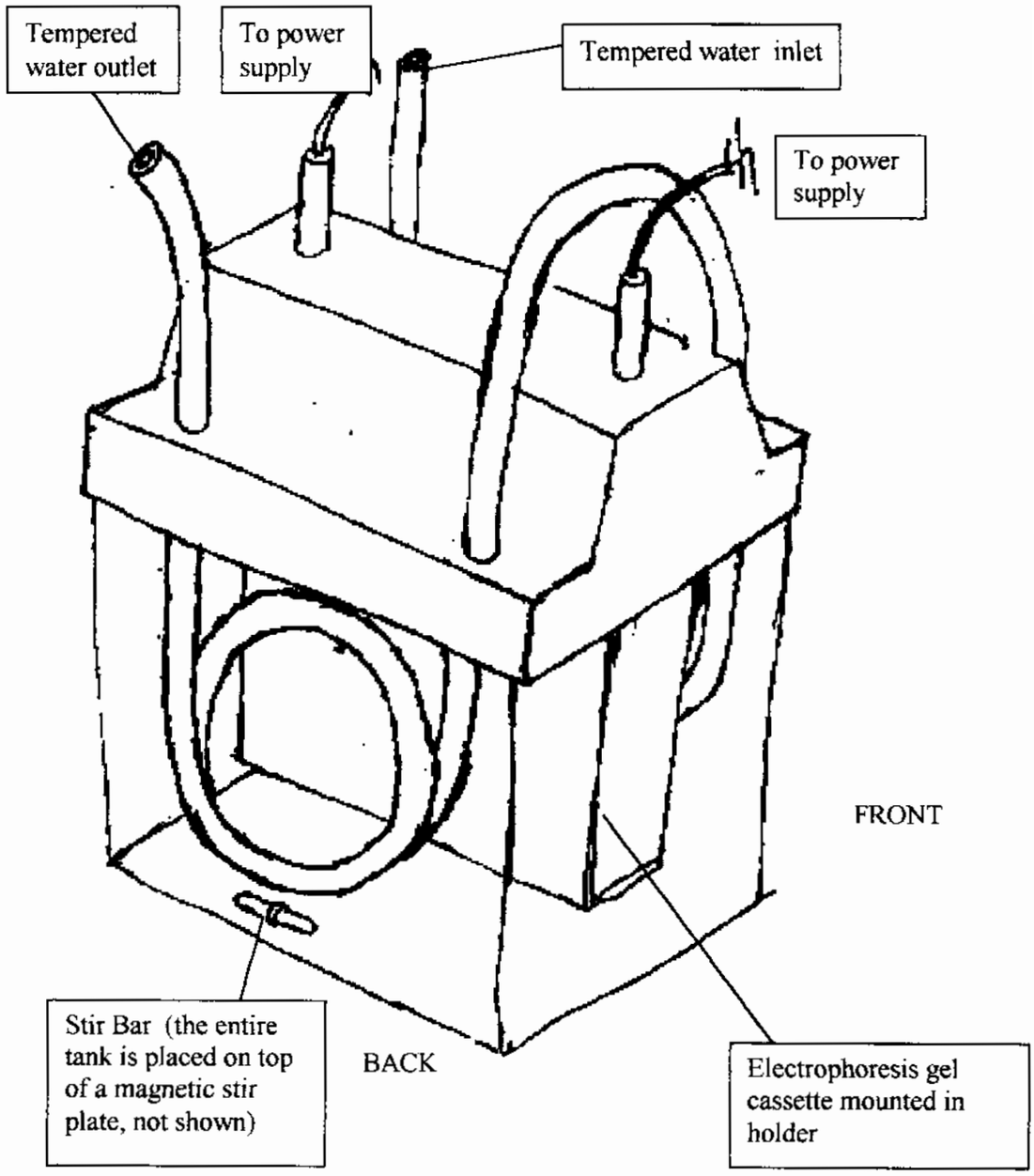


Figure 12: The modified Novex Excell II electrophoresis cell.

The Novex ExCell II electrophoresis tank will rapidly lose temperature over the course of a typical electrophoretic separation. Use of a "heating core", such as Tygon tubing connected to a high temperature water bath, will reduce this loss of temperature. While the modifications shown are specific to an ExCell II electrophoresis tank similar modifications can be made with tanks from other manufacturers, such as the Bio-Rad Protean II.



ExCell II tank, and similarly sized electrophoresis units such as the BioRad Protean II, is necessary to ensure that the electrophoresis temperature remains between 50 – 60 °C.[49] The tubing was connected to a Fisher Scientific Isotemp model 800 water bath that allowed external control of the chamber temperature through thermal transfer between the tubing and chamber. To ensure even buffer distribution throughout the tank, a magnetic stir bar was inserted inside of the electrophoresis tank, and the tank was placed on top of a magnetic stir-plate (Corning Scientific) set to medium rotation. The Novex electrophoresis unit however was limited to twenty samples, and maximum water bath temperatures of 80 °C due to limitations of the Tygon tubing. This restriction limited the maximum operational temperature inside of the modified ExCell II tank to 60 °C.[50] For large numbers of samples, or samples requiring greater resolution, a Hoeffer model SE600 electrophoresis unit equipped with a heating/cooling core was utilized.[51] The electrophoresis tank was connected to a Fisher Scientific Isotemp model 1016D water bath for heating and cooling operations. To ensure even buffer distribution, a magnetic stir bar was inserted inside of the electrophoresis tank, and the tank placed on top of a magnetic stir-plate (Corning Scientific) set to medium rotation. Electrical power was supplied by either an IsoLab, Inc. model CVR-500 power supply or an EC Apparatus Inc. model EC600 power supply. Acrylamide, bis-acrylamide and other electrophoresis reagents were obtained either from Sigma Chemical Co. or Bio-Rad, Inc. Acrylamide gradient gels were poured with a shop-built gradient maker consisting of two cylinders interconnected through a manually operated valve.

3.1.9 UV-Visible Spectrophotometer

Either a Cary model 300E Bio or Cary model 100E spectrophotometer from Varian, Inc. was used for all spectral measurements. Both instruments were interfaced to Windows-based computers with Cary/Varian WinUVBio version 2.0 software. Both spectrophotometers were equipped with a Peltier thermoelectric heating/cooling block, a multicell-transport device and a sample compartment with nitrogen-purge capability. Both machines utilized water baths manufactured by Varian for the Peltier heat sink. The machines were also equipped with external probe sets for direct temperature measurement inside of the sample cuvettes. The optical performance of each machine was regularly checked against calibrated neutral density filters.[52]

3.1.10 Lyophilizer

A Savant Instruments, Inc. model SVC-100H "SpeedVac" equipped with a Savant refrigerated vapor trap was used for all lyophilization procedures. The system was connected to a Precision Scientific high vacuum pump model DD90 (Fisher Scientific) or an Edwards model E2M5 two-stage (Thomas Scientific) vacuum pump.

3.1.11 Deprotection ("dip") Tank

A covered heating bath containing silicon oil (both from Fisher Scientific) was allowed to equilibrate between 50 – 55 °C at all times.[53] Immersed in the oil bath was a miniature "Lab-Jack" (Fisher Scientific) with a 500 ml beaker containing silicon oil at approximately the same solution level as the surrounding oil bath. The lab jack was utilized to adjust the oil level between the beaker and the silicon oil bath

3.1.12 Dialysis Equipment

Cellulose acetate dialysis membrane tubing with a 1000 molecular weight cutoff (MWCO) pore size, and dialysis tubing clips from Spectro-Por Industries, were used for all dialysis procedures. In addition, unwaxed and unflavored dental floss[54] (Johnson and Johnson, Inc.), magnetic stir-plates (Corning Scientific, Fisher Scientific, and others), and refrigerated "deli-cases" with pass-thru doors (LabRepCo, Inc.) were utilized.

3.1.13 Digital Photography Equipment

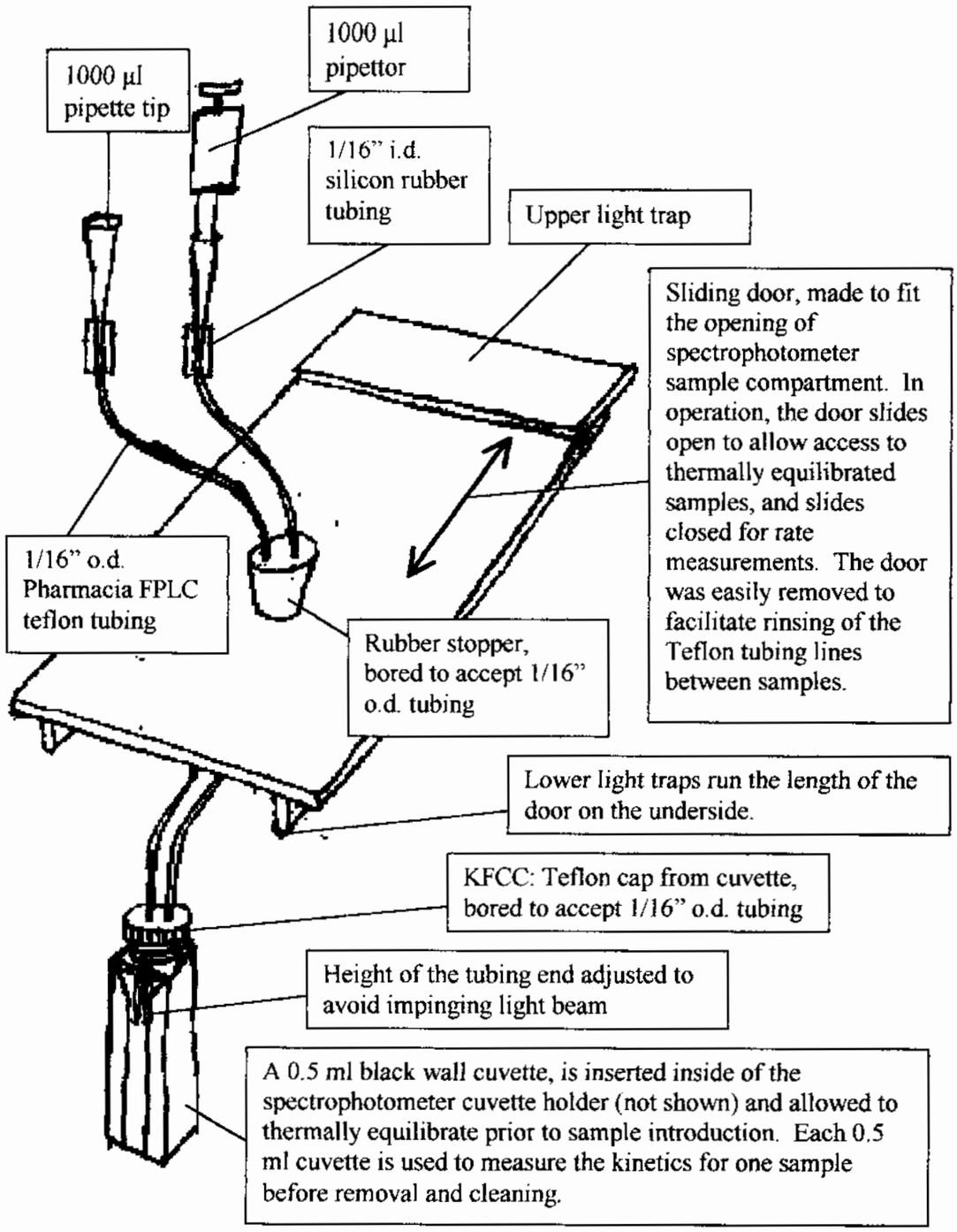
An AlphaImager digital camera, (Alpha Innotech Inc.) model 4912-2810 was mounted to a Polaroid MP-7 copy stand. The camera was directly interfaced to a Windows-based computer running AlphaEase version 3.3d software for control of the camera, image capture, densitometry ,and image manipulation. Included with this camera was a UV-cut-off filter, and a +2 close-up diopter lens enabling macrophotography. The light sources were either a UV-transilluminator, a fluorescent transilluminator, or overhead 500 W reflector floods. For all gel electrophoresis experiments in this project, a UV-transilluminator was used.

3.1.14 Kinetics Flow Cell

A 0.5 ml, long-range UV cuvette with a 10 mm path length and 2 mm path width (Helma Scientific) was modified to allow direct sample introduction without the need for removal of the cuvette from the spectrophotometer, or opening of the cuvette bay door (Figure 13). To accomplish this, two 1/16 inch o.d. teflon tubing lines were inserted through a teflon stopper designed (called the kinetics flow cell cap or KFCC) to fit the opening taper of the cuvette. The tubing openings were bored slightly oversized to allow easy

Figure 13: The kinetics flow cell cap (KFCC) and sliding door assembly.

The design of the sliding door and kinetics flow cell cap (KFCC) assembly. The penultimate end of the tubing has two short lengths of 1/16 inch i.d. silicon rubber tubing for insertion of a pipette tip. On one end, a 1000 µl pipette tip acts as an overflow reservoir. On the other end is a 1000 µl pipettor used to introduce sample solution and mix the contents of the cuvette. (Note: Not drawn to scale)



insertion of the teflon tubing. The length of the tubing below the bottom of the stopper was approximately 1 cm – well above the lightpath of the spectrophotometer. Above the teflon stopper, the two teflon tubing lines were inserted through a rubber stopper that formed the light seal of a shop-made secondary spectrophotometer door. The door design incorporated light traps along the exposed edges of the door (to eliminate stray light), a hole positioned over position 3 of the cuvette cell holder (to ensure the tubing lengths were kept to a minimum as well as to enable the holder to move front to back) and sufficient length to enable the door to be slid open for access to samples in the equilibration cuvettes and closed during measurements (Figure 13). The door was removable without any extra tools to facilitate rinsing of the Teflon lines between samples and removal of the cuvette. At the conclusion of one measurement, the measurement cuvette was removed, another cuvette of similar dimensions was inserted in the spectrophotometer, and the cleaned KFCC assembly was inserted in the cuvette.

3.2 Definition of Common Solutions and Reagents

3.2.1 SHU Denaturing Gel Sample Buffer (for TBE-based PAGE)

1X Tris-Borate-EDTA (TBE defined below, 3.2.4) solution pH 8.0, 90% v/v formamide, 0.1% w/v bromophenol blue, 0.1% w/v xylene cyanol.[5556]

3.2.2 SHU Denaturing Gel Sample Buffer (for TAE-based PAGE)

1X Tris-Acetate-EDTA solution (TAE, defined below, 3.2.5) pH 8.0, 90% v/v formamide, 0.1% w/v bromophenol blue, 0.1% w/v xylene cyanol.[5756]

3.2.3 SHU Native Gel Sample Buffer

5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1mM EDTA, pH 7.0, 100 mM NaCl, 50% v/v glycerol, 0.1% w/v bromophenol blue, 0.1% w/v xylene cyanol.

3.2.4 Tris-Borate-EDTA (TBE)

45 mM tris-borate, 1 mM EDTA, pH 8.0 with boric acid. To prepare one liter of solution, add 5.4 g Tris base, 2.75 g boric acid, 2.0 ml of 0.5 M EDTA, to 750 ml distilled, deionized water, then adjust pH to 8.0 with boric acid if necessary. Adjust volume one liter with d.d. water

3.2.5 Tris-Acetate-EDTA (TAE)

40 mM tris-acetate, 1 mM EDTA, pH 8.0 with acetic acid. To prepare one liter of solution, add 4.84 g Tris base, 1.14 ml of glacial acetic acid, and 2.0 ml of 0.5 M EDTA to 750 ml of distilled deionized water. Adjust pH to 8.0 with acetic acid if necessary, then adjust volume to one liter with water.

3.2.6 50X TAE (50X Concentrated TAE Stock)

2.0 M tris-acetate, 50 mM EDTA, pH 8.0 with acetic acid.

3.2.7 PBS (Phosphate Buffered Saline)

5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1mM EDTA, pH 7.0, 100 mM NaCl.

3.2.8 PBS (Phosphate Buffer Solution)

5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1mM EDTA, pH 7.0.

3.2.9 10X PBS (for purposes of DNA sample preparation)

1.0 mM EDTA, 50 mM NaH_2PO_4, 50 mM Na_2HPO_4, pH 7.0.

3.2.10 10X NaCl (for purposes of DNA sample preparation)

1.0 M NaCl.

3.2.11 Acrylamide Stock Solution

40% w/v acrylamide in water composed of 19:1 w/v of acrylamide and bis-acrylamide (to 60 ml of water add 38 g of acrylamide, 2 g of bisacrylamide and dissolve. Adjust volume to 100 ml with water)

3.2.12 0.1 M TEA/A

0.1 M triethylammonium acetate, (TEAA) pH 7.0. To prepare this solution, the calculated quantity of triethylamine is added to water, and the pH is adjusted to 7.0 with acetic acid.

3.2.13 2.0 M TEA/A

2.0 M triethylammonium acetate pH 7.0. To prepare this solution, the calculated quantity of triethylamine is added to water, and the pH is adjusted to 7.0 with acetic acid.

3.3 Oligomer Synthesis and Initial Characterization of Oligomers

3.3.1 Synthesizer Optimization Measurement Procedures

As in most chemical reactions, the reactant concentration, the amount of each reactant participating in the chemical reaction (stoichiometry and limiting reagents) the length of the reaction time, and the reaction temperature are critical to obtaining high product

yields. With respect to the model 380B DNA synthesizer the amount of each reagent applied to the column determines, to a great extent, the quality of the synthesis reactions. If too little reagent is applied to the column, the cyclic synthesis reactions do not proceed to completion, the step-wise yields decrease, and the amount of failure sequences increase.[58] The amount of reagent applied to the controlled pore glass CPG column is defined by three operational parameters – reagent concentration, flow rate and flow time. The reagent concentration is a fixed parameter determined by the ABI specifications for this synthesizer and the reaction stoichiometry. The flow rate is also fixed parameter determined by the internal diameter (i.d.) of the delivery line.[59] The i.d. of the delivery lines will vary according to the reagent. The acetonitrile delivery lines are the largest, and afford the highest flow rates to the column, whereas the phopshoramidite lines are among the smallest on the synthesizer. Not surprisingly, the phosphoramidite lines are also the most prone to clogging. Partial clogs can alter the internal diameter of the reagent delivery line and reduce the overall flow rate. To ensure optimal performance, the flow rates from all reagent delivery lines were routinely tested with acetonitrile and compared to the ABI flow-rate specifications for each delivery line (Table 5).[60] Those reagent supply lines with out-of-spec delivery times were noted, and appropriate corrections were made to the reagent delivery times within the ABI synthesis cycle.

Other factors, such as the age of the reagents on the synthesizer, and the relative humidity in the laboratory, were found to affect synthesis yields. The age of the reagent decreased the available concentration and increased the concentration of degradates that can interfere with the synthesis reactions. Humidity in the laboratory also degraded synthesis

Table 5: Flow rate specifications for the ABI model 380B DNA synthesizer.

The flow rate specifications listed are for acetonitrile in all positions. Do not attempt to measure the flow rates with other solutions, such as the synthesis chemistry, because the differing viscosities of these solutions will affect the measurement. The flow rates listed are for the 380B synthesizer with version 1.34 software (important).

The flow rates were measured as follows AT THE COLUMN INLET:

1. In manual mode, program the synthesizer to deliver acetonitrile to the column.
2. Allow the acetonitrile to flow for 30 seconds prior to measurement to release any trapped bubbles of argon in the delivery line.
3. Collect the acetonitrile for the specified period of time or volume. If using the time/volume approach use a 1 ml (or 5ml if you have the patience) volumetric flask.
4. The ABI specifications appear below.

Reagent	Position	CH_3CN flow gm / 60 sec	CH_3CN flow ml / 60 sec	CH_3CN flow sec / ml
A	1	0.57-0.60	1.45-1.55	0.64-0.69
C	2	0.57-0.60	1.45-1.55	0.64-0.69
G	3	0.57-0.60	1.45-1.55	0.64-0.69
T	4	0.57-0.60	1.45-1.55	0.64-0.69
open	5	0.57-0.60	1.45-1.55	0.64-0.69
open	6	0.57-0.60	1.45-1.55	0.64-0.69
open	7	0.57-0.60	1.45-1.55	0.64-0.69
open	8	0.57-0.60	1.45-1.55	0.64-0.69
Activator	9	0.57-0.60	1.45-1.55	0.64-0.69
NH_4OH	10	1.53-1.77	3.90-4.50	0.22-0.26
Cap A	11	0.63-0.79	1.60-2.00	0.5-0.63
Cap B	12	0.63-0.79	1.60-2.00	0.5-0.63
CH_3CN	13	0.65-0.71	1.65-1.8 (see note 1)	0.55-0.61
Deblock	14	0.65-0.71	1.65-1.8	0.55-0.61
Oxidizer	15	0.65-0.71	1.65-1.8	0.55-0.61
CH_3CN	16	0.65-0.71	1.65-1.8 (see note 1)	0.55-0.61
CH_3CN	18	0.65-0.71	1.65-1.8	0.55-0.61

NOTES:

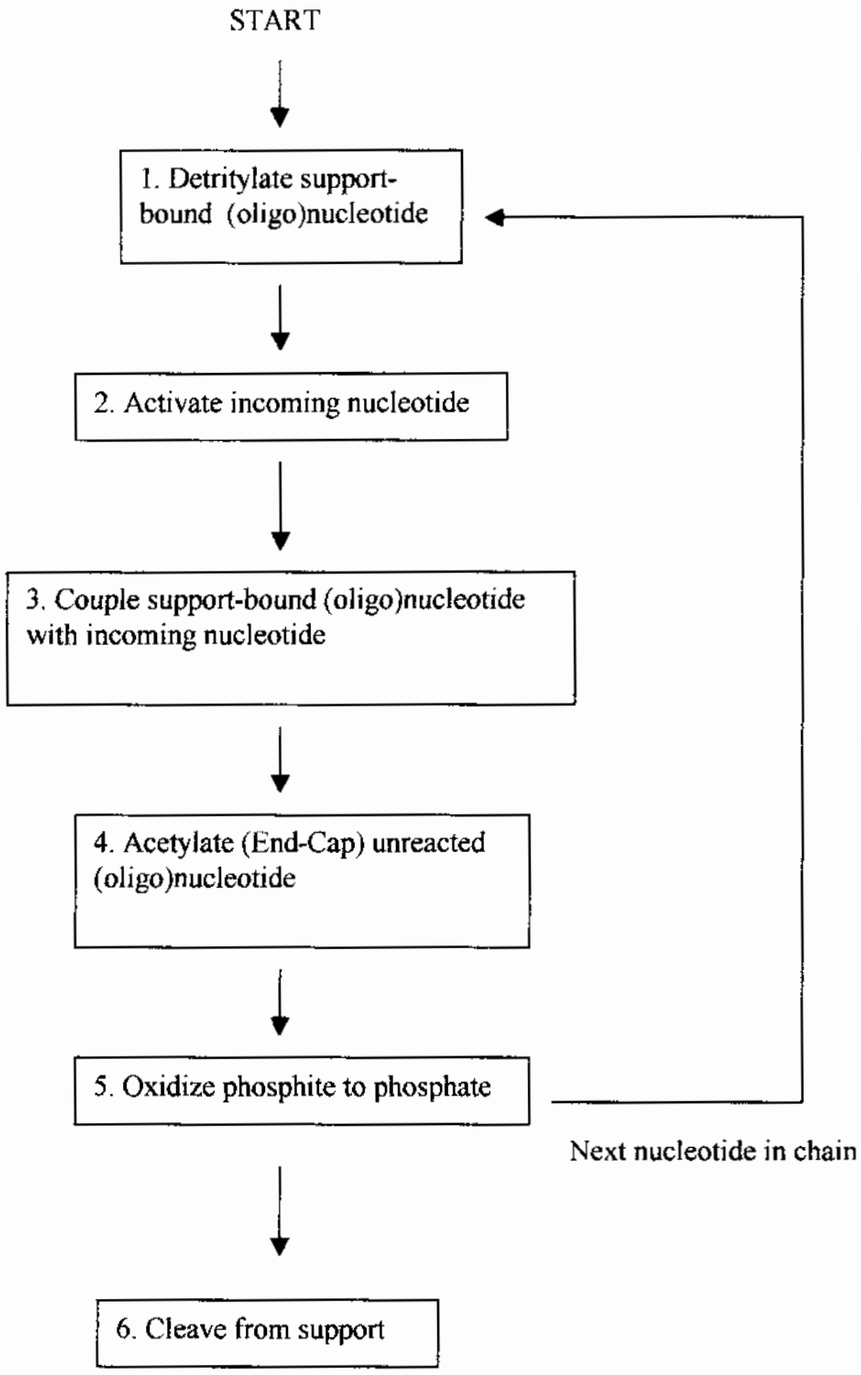
1. ABI does not list a flow-specification for acetonitrile in positions 13 or 16. I can only assume the flow rate for positions 13 and 16 is the same as specified for the acetonitrile in position 18.

yields by providing a competing nucleophile during the coupling reaction. The seals on the ABI synthesizer, while good, were not perfect. With age and humidity in mind, the phophoramidites were not used beyond two weeks on the synthesizer regardless of the remaining solution level. The activator, and capping reagents were not used beyond one month once mounted on the synthesizer. The oxidizer, deblocking and acetonitrile solutions were not used beyond two months on the synthesizer.[61] The shelf-life of all reagents in unopened bottles were longer than those listed above, depending upon the storage conditions. Acetonitrile, transferred from an Aldrich SureSeal bottle and stored in amber glass bottles under dessication was not used beyond 6 months. Acetonitrile in an unopened SureSeal bottle was used without regard to reagent age. Lyophilized phosphoramidites, stored under dessication were used within six months, or cautiously used beyond 6 months. All other reagents were stored as unopened bottles in the dark under ambient conditions and used without regard to age unless otherwise indicated.

3.3.2 Schematic Overview of Oligonucleotide Synthesis

All synthesis procedures were performed on a 1μmole scale and produced on average between 1000 to 1500 μg of crude DNA per synthesis. All synthesis reactions occurred on a controlled pore glass (CPG) solid support to allow efficient removal of by-products and unreacted reagents. All oligomers were synthesized as 5'-terminal dimethoxytrityl (DMT-on) oligomers using the phosphoramidite method of Caruthers (Figure 14).[62] Synthesis began at the 3' terminus of the oligonucleotide and preceded in the 3'- to 5'-direction to take advantage of the reactive primary alcohol in the 5' position. The first nucleoside in the sequence is covalently bonded to the CPG via a succinyl ester linker. The first step in synthesis is cleavage of the dimethoxytrityl (DMT) moiety from the 5'

Figure 14: Overview of the phosphoramidite synthesis procedure.



end of the bound nucleotide under acidic conditions. In the second and third steps, which occurred simultaneously on the synthesizer, the incoming nucleoside base was activated with tetrazole and covalently coupled to the deprotected nucleoside on the CPG support. The net result of successful incorporation is extension of the oligomeric chain by one nucleotide, and incorporation of a DMT group on the 5'-terminus of the DNA. The reaction of the incoming nucleoside with the bound nucleoside, however, is not 100% efficient and resulted in a small population of failure sequences.[63] In the fourth step the failure sequences are "end-capped" or acetylated with acetic anhydride to prevent further reaction with the next synthesis cycle. Failure to "end-cap" the failure sequences would result in a population of N-1, N-2, etc., deletion sequences. In the fifth step the backbone phosphite is oxidized to a phosphate. After oxidation the synthesis cycle either begins again for the next nucleotide in the sequence or is stopped and the oligonucleotide cleaved from the solid support by base hydrolysis.

3.3.3 Chemical Details of DNA Synthesis by the Phosphoramidite Approach

3.3.3.1 Controlled Pore Glass Column Selection

The identity of the synthesis column was determined by the 3' nucleotide of the sequence to be synthesized - for all (5'-(CNG)n-3') triplet repeat sequences, a guanine controlled pore glass (G-CPG) column was used. The porosity of the controlled pore glass (CPG) column was critical to obtaining good synthesis yields. Long oligonucleotides required larger pore sizes to allow incoming reagents enough space to diffuse and react with the support-bound oligomer. Therefore, 500 Å pore-size CPG columns were used for all oligomers shorter than 50 nucleotides in length, and 1000 Å pore-size CPG columns were

used for oligomers greater than 50 nucleotides in length. Covalently attached to the CPG was a succinyl ester linker coupled to a G-deoxyribosyl moiety protected in the 5'-position by a dimethoxytrityl group (Figure 15). Depending upon the nature of the structure, the exocyclic ring nitrogens, were protected with either an isobutyryl or benzoyl moieties (Figure 16).

3.3.3.2 Dimethoxytrityl Group Cleavage Reaction

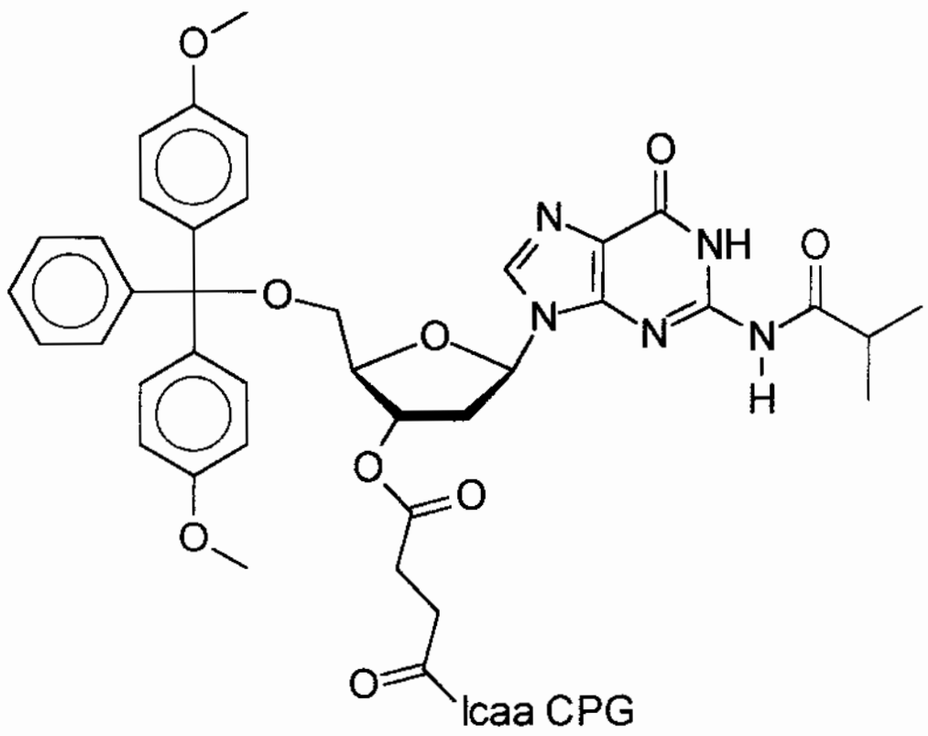
The 5'-OH group of the deoxyribose sugar was protected to prevent side reactions, such as self-polymerization, from occurring. However, prior to the coupling reaction, the DMT group had to be removed to provide a reactive site for the incoming phosphoramidite. The dimethoxytrityl (DMT) protecting group was readily cleaved from the growing oligonucleotide with 3% trichloroacetic acid (TCA) in dichloromethane (DCM). The reaction mechanism follows that of an ether hydrolysis reaction, where the nucleophilic ether oxygen attacks the acid proton to form an oxonium ion which undergoes an S_N1 cleavage reaction to form the desired oligonucleotide (Figure 17).

3.3.3.3 Coupling Reaction

The coupling reaction is sensitive to both air and moisture. Water, in the form of moisture, acts as a competing nucleophile in the coupling reaction. The column was flushed with argon, then rinsed with low moisture acetonitrile prior to the coupling reaction. The deprotected oligonucleotide was coupled to the next nucleoside base in the growing DNA chain by first activating the incoming nucleoside with 0.45 M tetrazole in acetonitrile. Tetrazole protonated the diisopropylamine tertiary nitrogen making the protonated form, diisopropylaminohydride, a good leaving group. The nucleophilic

Figure 15: The G-nucleoside with a succinyl linkage and DMT-protecting group.

The guanine nucleoside, with a dimethoxytrityl (DMT) group (shown in red), the isobutyryl protecting group (shown inmagenta) and the succinyl linker to the controlled pore glass (shown in blue) is shown below. The DMT group protects the 5'-oxygen from undergoing reaction with nucleophiles.



(lcaa CPG = long chain alkylamino Controlled Pore Glass)

Figure 16: The exocylcic amine protecting groups.

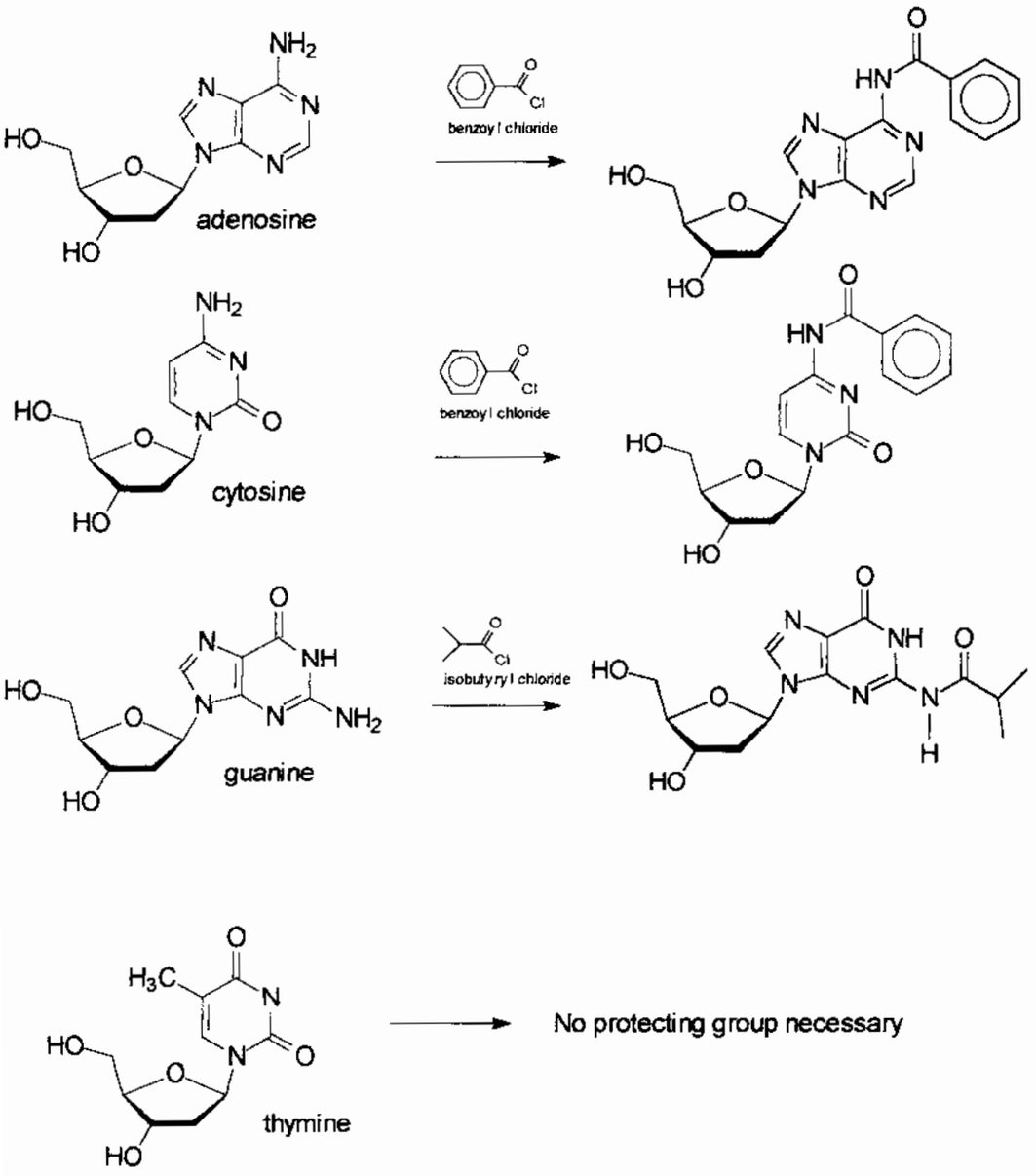
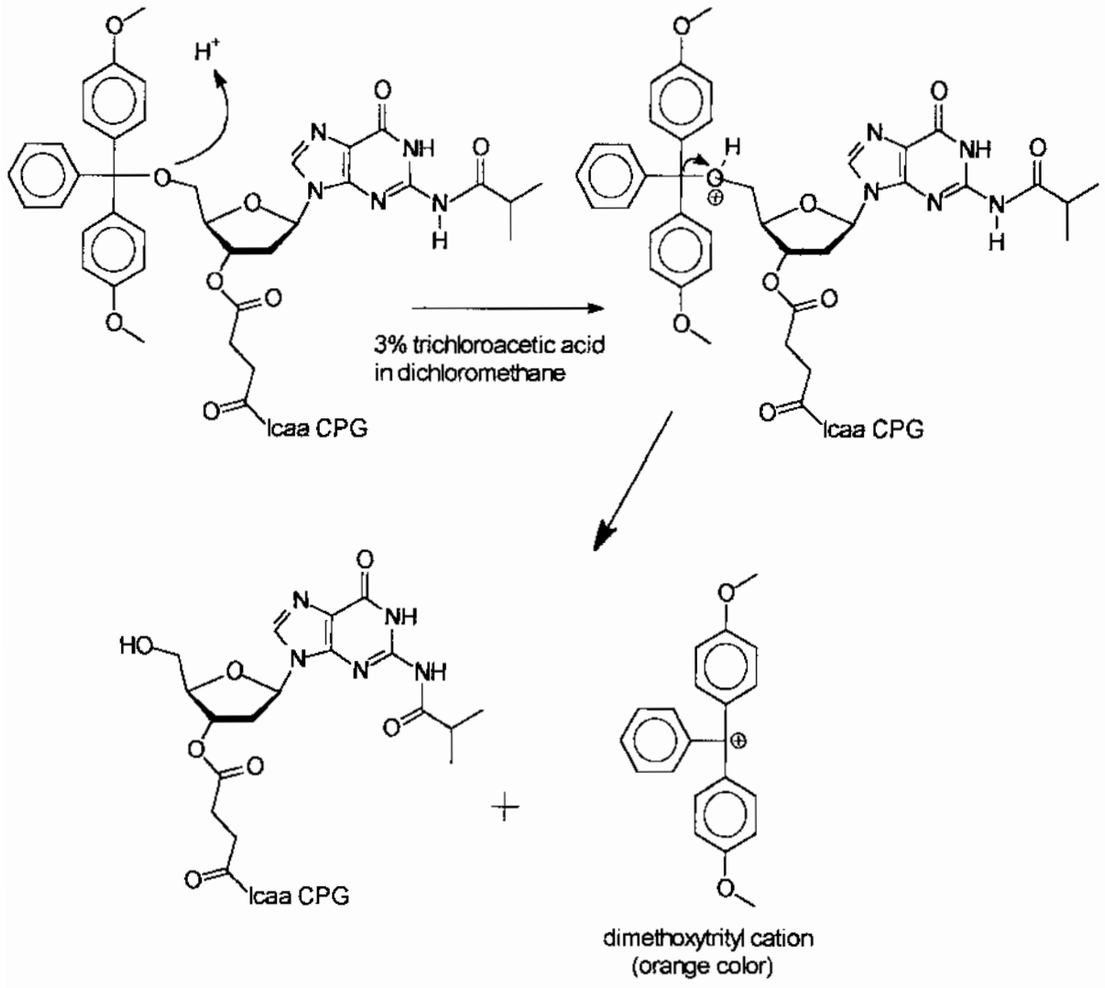


Figure 17: The reaction mechanism of DMT hydrolysis under acidic conditions.

As shown in the figure below, the DMT hydrolysis follows a reaction scheme typical of ether hydrolysis reactions.



hydroxyl group of the support-bound mononucleotide displaced the protonated imidazole by nucleophilic attack of the phosphorus (3+) resulting in cleavage at the phosphorus-nitrogen bond (Figure 18).

3.3.3.4 End-Capping Reaction

Each coupling reaction resulted in a small subpopulation (about 0.5 - 1%) of failure sequences that had to be acylated ("end-capped") at the 5'-hydroxyl to prevent further reactions from occurring at this site. Capping of the unreacted oligomers not only produced higher yields, but also allowed rapid purification at the end of the synthesis. The exposed 5'-hydroxyl of the failure sequence oligonucleotides was acylated by reaction with acetic anhydride and 10% methyl imidazole in tetrahydrofuran. The mechanism proceeds via a nucleophilic addition/elimination reaction. The hydroxyl oxygen attacks the carbonyl of acetic anhydride, forming a tetrahedral intermediate that later undergoes an elimination reaction to produce acetate ion and an acetylated nucleobase. The acetate ion then reacts with methylimidazole in an electrophilic addition to produce 1-methyl-4-acetoimidazole (Figure 19).

3.3.3.5 Oxidation Reaction

To prevent alkaline degradation of the growing oligonucleotide chain, it is necessary to oxidize the phosphite moiety to phosphates. Oxidation of the backbone phosphite group was accomplished by reaction of the oligonucleotide chain with 0.1 M iodine in water, tetrahydrofuran, and pyridine. The reaction proceeded by an oxidation-reduction

Figure 18: The β-cyanoethylphosphoramidite activation and coupling reactions.

Coupling between the incoming β-cyanoethylphosphoramidite , and the CPG-bound mononucleotide occurs after protonation of the tertiary amine on the phosphoramidite.

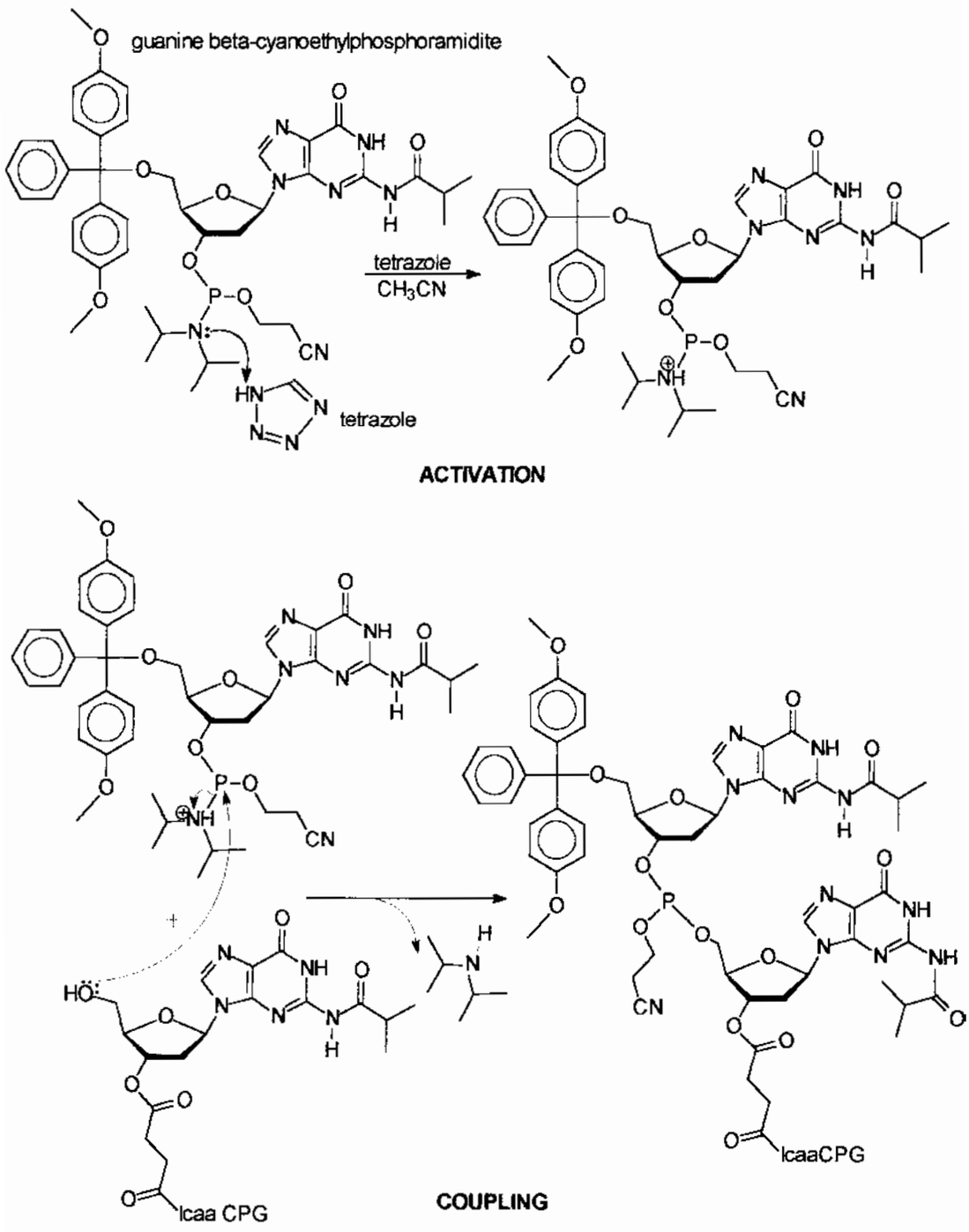
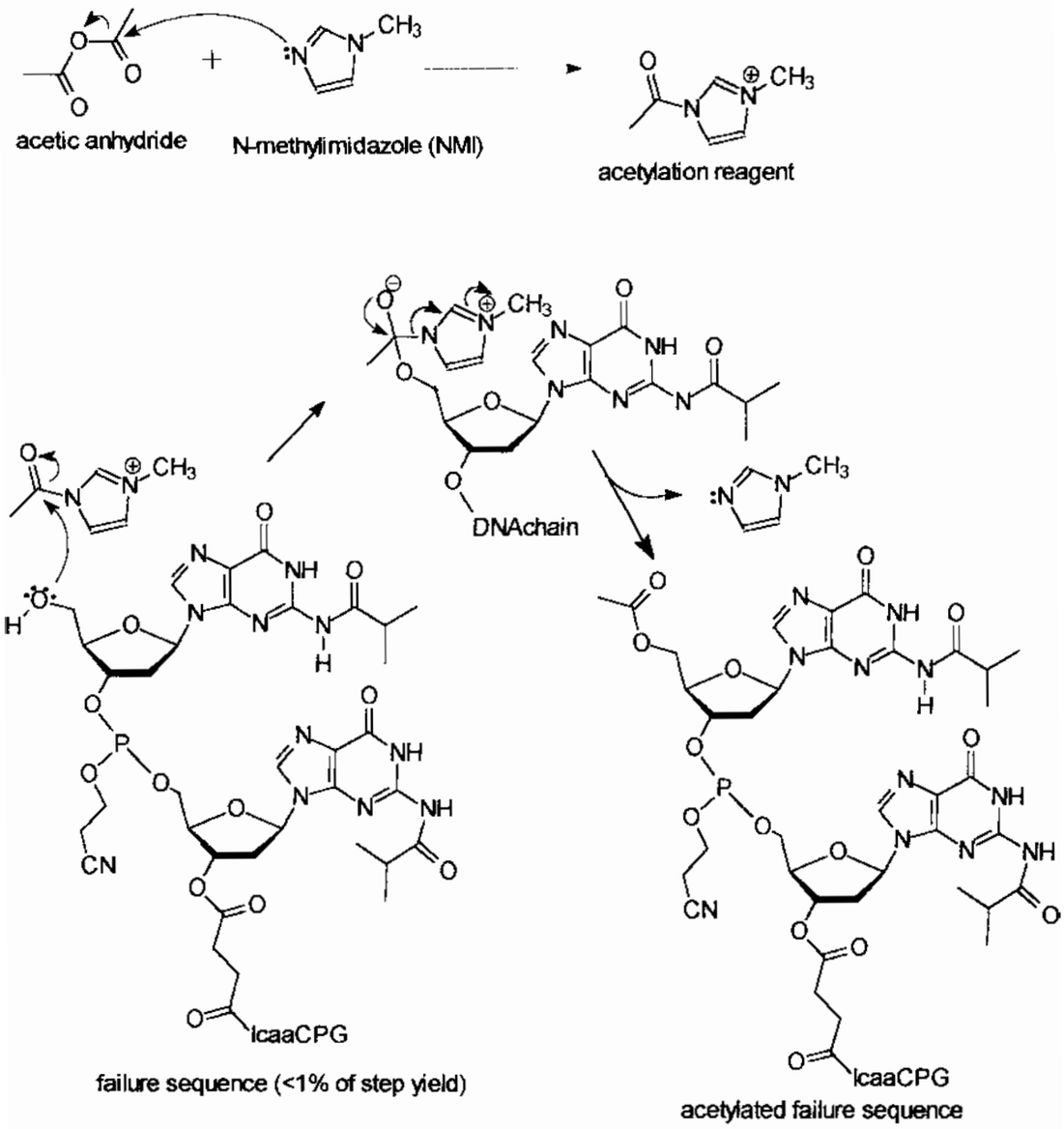


Figure 19: The acetylation or "end-capping" reaction.

The acetylation or "end-capping" of the failure sequences begins with the preparation of 2-acetylmethylimidazole. The acetylation reaction then proceeds through an addition/elimination mechanism.



mechanism, where the phosphite (phosphorus III cation) was oxidized to a phosphate (phosphorus V cation) by iodine (Figure 20) with water as the oxygen donor.

3.3.3.6 Oligonucleotide Cleavage from CPG

On conclusion of synthesis the oligonucleotide contained, besides the primary sequence of nucleobases, a 5'-terminal DMT protecting group, various protecting groups on the exocylclic amines, and a linkage to the CPG via a succinate ester. All of these groups had to be removed prior to experimental use. Cleavage of the DMT-on oligonucleotide from the solid support was accomplished on the synthesizer at room temperature with concentrated ammonium hydroxide. The succinate ester is base-labile, and under alkaline conditions will readily hydrolyze, via a base hydrolysis mechanism (Figure 21). The cleaved oligonucleotide was collected into a pyrex culture tube coated, on the outside, with a plastic safety layer. The coated collection tubes prevent catastrophic loss of the synthesized DNA should breakage occur.

3.3.3.7 Deprotection Reaction

The exocyclic amine protecting groups remain with the oligonucleotide after cleavage from the CPG solid support and require removal before any experimental procedures can be executed with the oligomer. The exocyclic protecting groups are base-labile, and undergo base-hydrolysis at elevated temperatures. To 2 ml of crude oligonucleotide solution (already in ammonium hydroxide) approximately 2-3 ml of fresh ammonium hydroxide was added to ensure a fully alkaline state.[64] The culture tube was tightly capped, wrapped in Parafilm and immersed in a 50 - 55 °C silicon oil bath for 18 hours.[65] Immersed in the oil bath was a miniature "lab-jack" (Fisher Scientific) with a beaker

Figure 20: Oxidation of the phosphite to phosphate with iodine.

The phophorus is oxidized from a 3^{+} oxidation state (phosphite) to a 5^{+} oxidation state (phosphate).

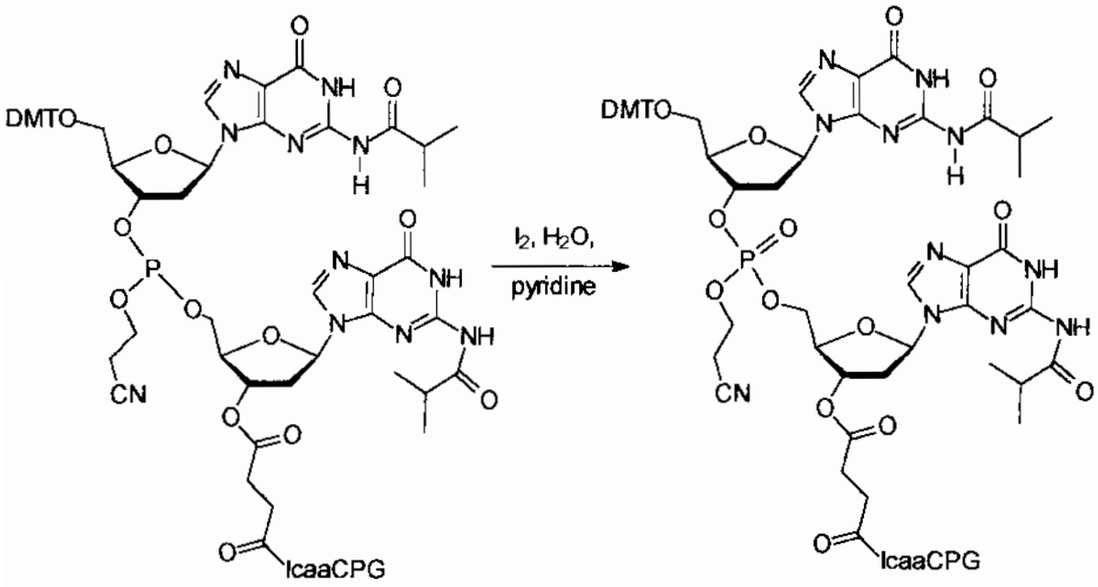
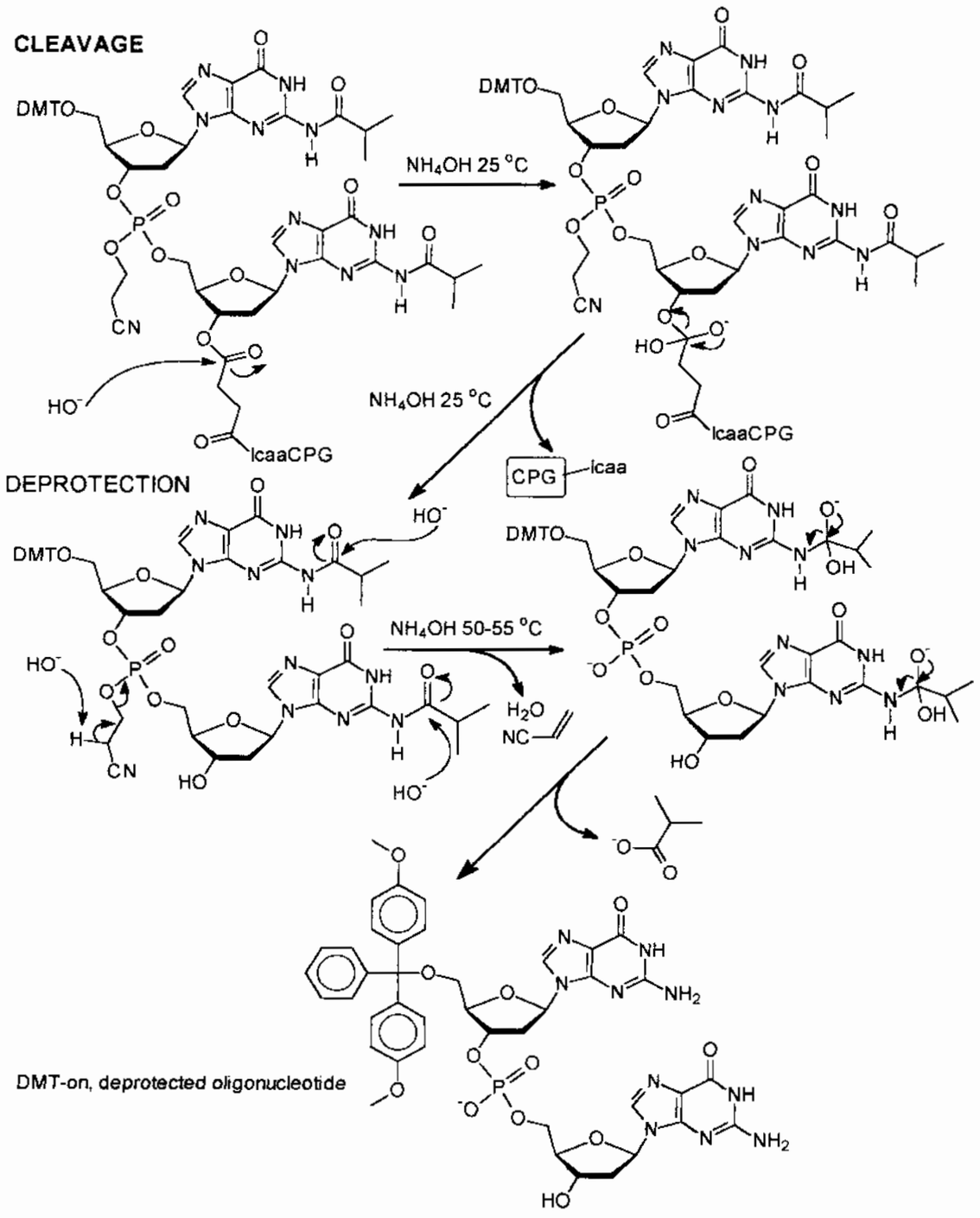


Figure 21: Cleavage of the oligonucleotide from the solid support and deprotection.

In the figure below reactants are shown on the left, and intermediates on the right. The product of the cleavage reaction is the reactant for the deprotection reaction.



containing silicon oil at the same solution level and temperature as the surrounding oil bath. Deprotection of multiple oligonucletide samples was accomplished by placing the tightly capped tubes within the beaker. The solution level of silicon oil within the beaker was adjusted to match that of the oligomer solutions by adding or removing oil in the beaker.[66] Finally, the solution level of the beaker was adjusted, using the lab-jack to match that of the surrounding oil bath.[67]

3.3.3.8 Trityl Assay

Monitoring of the overall and step-wise trityl yields provided diagnostic information about the synthesis step efficiency, the overall yield, and oligomer length in real-time. After each DMT-on synthesis the number of trityl peaks were counted and any oligomer with less than the full length number of peaks, defined as (n-1) where n = oligomer length, was immediately discarded.[68] Any oligomer with a trityl assay that gradually declined and did not recover, indicating a loss of step efficiency, high levels of failure sequences and an overall poor yield was immediately discarded [69] Those oligomers with an acceptable trityl assay, defined as the correct number of peaks and the correct peak heights throughout the entire synthesis were retained for further purification by RPHPLC or PolyPakII (Figure 22). The overall yield was calculated as the measurement of the final cycle peak height divided by the cycle 1 peak height (Equation 2). The step yield was approximated as the "n+1" peak height divided by the "n" peak height (Equation 3) or more accurately based upon the overall synthesis yield (Equation 4).

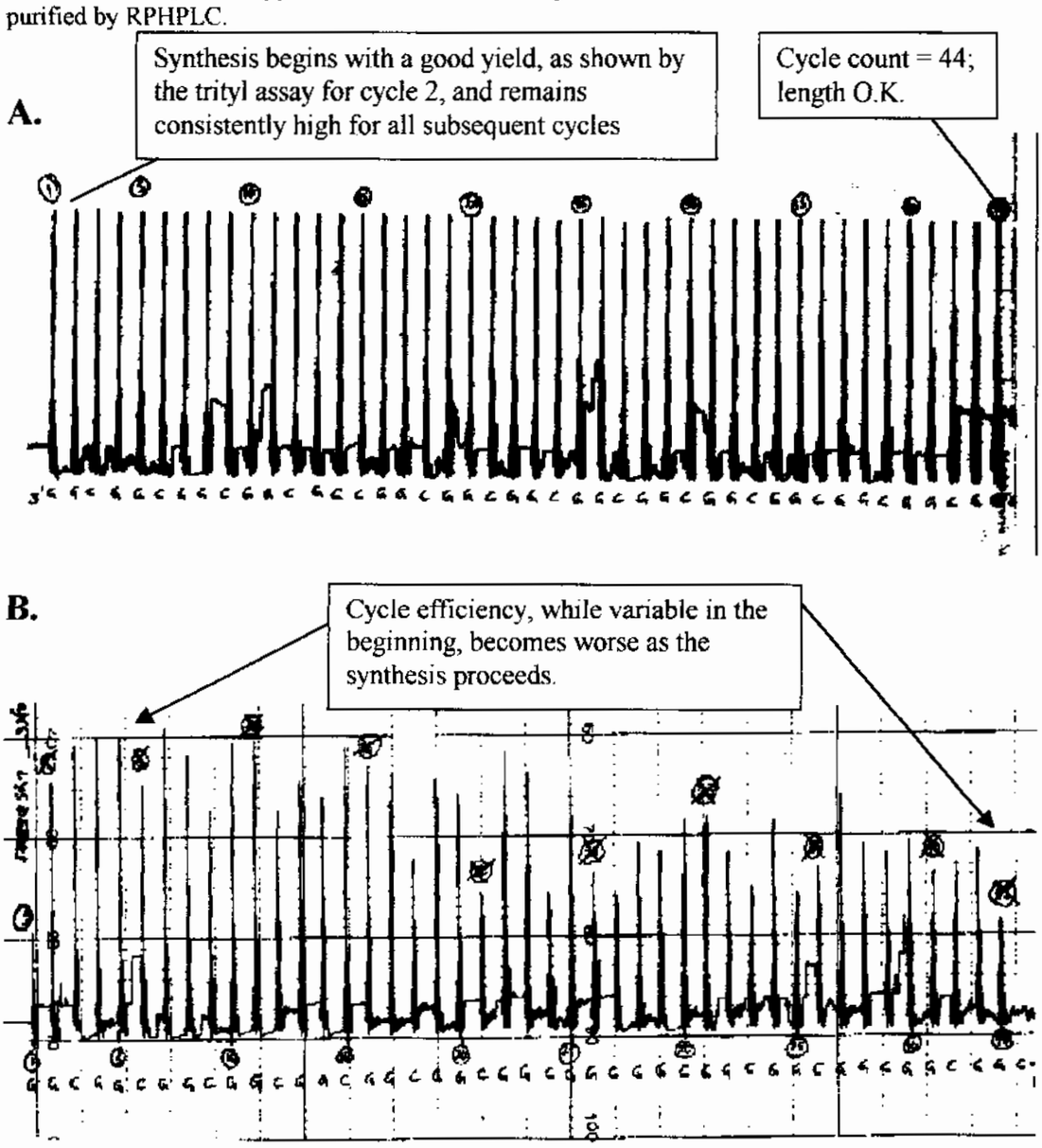
Equation 2: Overall Yield = (Peak Height $_{last}$/ Peak Height $_{first}$)*(100%)

Equation 3: Step Yield = (Peak Height $_{n+1}$ / Peak Height $_{n}$)*(100%) where n = coupling number

Equation 4: Step Yield = (Overall Yield)$^{1/n-1}$ where n = number of couplings

Figure 22: The trityl assays for two oligonucleotides.

A. A trityl assay for an acceptable oligonucleotide synthesis. All peaks heights are equivalent from N = 1 to N = 45, 44 peaks observed, with no deviation in peak height from cycle to cycle. This oligonucleotide is a good candidate for purification with a PolPak II cartridge column. **B.** A trityl assay of an unacceptable oligonucleotide (aka "trash-mer"). Note the variable peak height that begins around N =5, become progressively worse as synthesis proceeds. By cycle 44, the overall efficiency has dropped to about 50%. This oligonucleotide, if used at all, would be purified by RPHPLC.



3.3.3.9 Initial Characterization of the Crude Oligomer Yield

All oligomers were initially characterized post-synthesis by measuring the DNA concentration with a spectral scan from 400 nm to 200 nm. In a typical assay, 10 µl of crude oligomer was diluted 1:50 with 490 µl of phosphate buffered saline (PBS, 5 mM monobasic sodium phosphate, 5 mM dibasic sodium phosphate, 0.1 mM EDTA, pH7.4 , 100 mM sodium chloride). The wavelength of maximum absorbance and the absorbance at 280 nm were recorded according to the absorbance assays described in detail later. The DNA concentration was calculated with the dilution corrected absorbance value at 260 nm (Equation 5), using the relationships shown below (Equation 6, Equation 7).

Equation 5: **dilution = (total, prepared sample volume)/(aliquot volume)**

Equation 6: **$1A_{260}$ oligomer = 27 µg DNA**

Equation 7: **(A_{260})(dilution) / 0.027 = µg DNA/ml**

The total amount of DNA synthesized, in the form of failure sequences and full-length DMT oligomer was calculated by multiplying the DNA concentration by the volume of the crude oligomer solution (Equation 8).

Equation 8: **(µg DNA/ml)(ml solution) = µg total DNA**

A typical 1-µmole oligomer synthesis yielded from 500 to 800 µg of crude 12-mer DNA and from 800 to 1500 µg of crude 45-mer DNA. Oligomer solutions with DNA yields lower than 300 µg, indicating low yield and possibly high levels of failure sequences, were discarded and synthesis reattempted.

3.4 Oligomer Purification

3.4.1 Purification Scheme Overview and Flow Chart

The purification schemes adopted depended, in large part, upon the quality of the synthesized oligonucleotide as indicated by the trityl assay and the initial absorbance scan. Regardless of the purification method adopted, the decision to use an oligonucleotide depended on the purity and quality of the final product as indicated by a combination of RPHPLC, AXHPLC or Urea-PAGE electrophoresis. The definitive assay used for oligomer purity and quality was denaturing Urea-PAGE analysis. The purification scheme is shown schematically below (Figure 23).

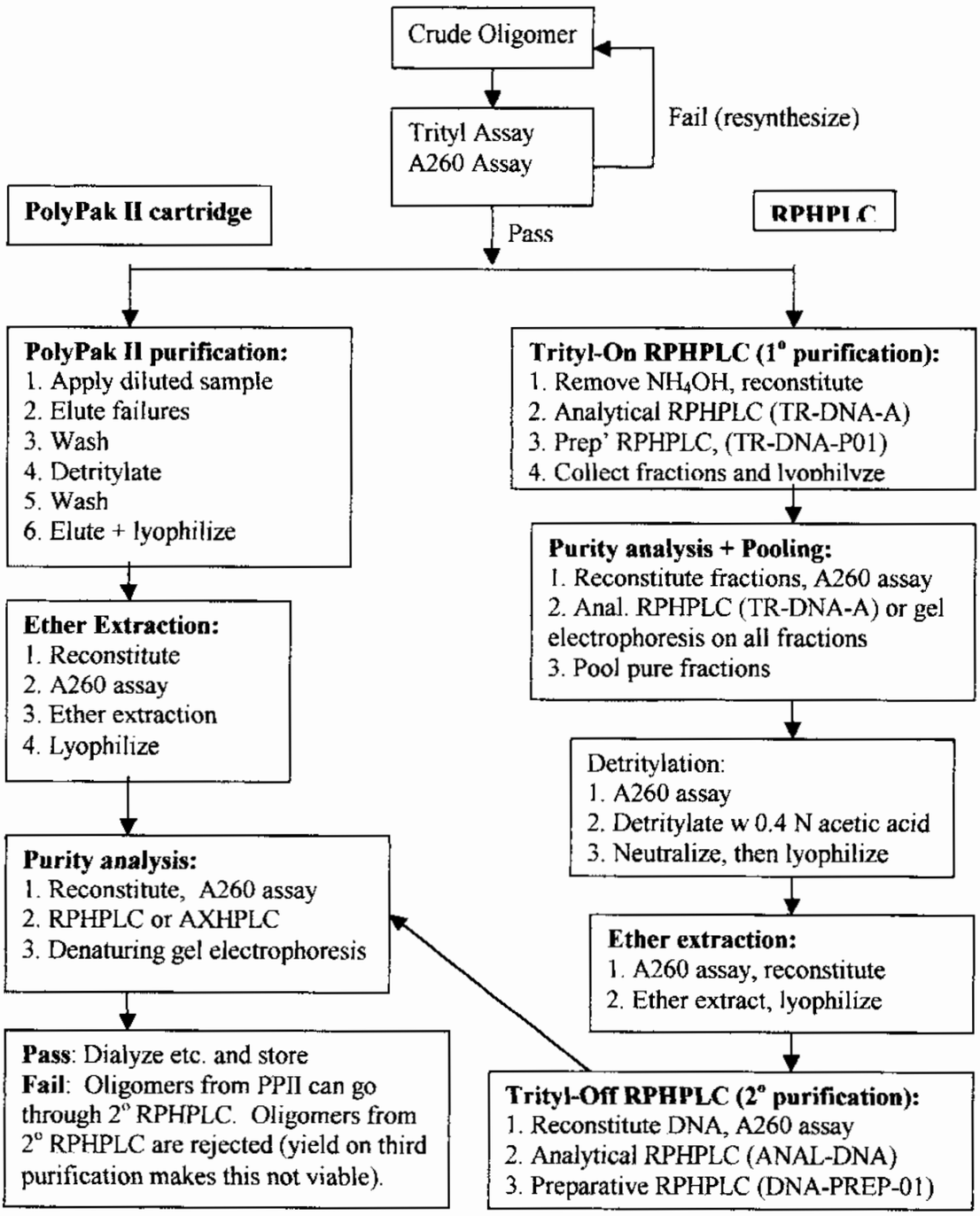
3.4.2 PolyPak II Purification, Manual Mode

3.4.2.1 Why PolyPakII, and When To Use a PPII column

PolyPakII columns provide speed, with processing times of a few hours instead of days, high capacity, acceptable selectivity, and the ability to prepare several oligomers simultaneously. In addition, these columns produced higher yields than RPHPLC because fewer lyophilization steps were required from crude to purified oligomer. These columns also allow the end-user to perform buffer exchanges and DNA concentration procedures, while the oligomer is still bound to the solid-support. These columns, however, were not applied for all purification situations. Oligomers from optimal syntheses, consisting of the correct length and high yield were ideal candidates for PPII purification procedures. In contrast, oligomers from less than optimal synthesis

Figure 23: An overview of the purification process.

Purification by PolyPak II cartridge column chromatography is suitable for pure oligomers that are needed in high yield. Lower yield, high purity oligomers can be obtained by RPHPLC separation, or a combination of both techniques. Regardless of the purification method selected, the oligomers must pass rigorous purity criteria as determined by denaturing polyacrylamide gel electrophoresis and other assays.



reactions, where there was a high probability of failure sequences, were not good purification candidates due to the selectivity and/or resolution of the PPII column. For this reason, oligomers from syntheses with variable yields, were not processed with PPII columns. In final analysis the decision to use an oligomer for experimental purposes was determined on the basis of analytical RPHPLC, AXHPLC, or, in a definitive manner, by analytical gel electrophoresis.

3.4.2.2 Principles of Operation

Automated synthesis is normally carried out in the 3' to 5' direction with a dimethoxytrityl (DMT) group on the 5'terminus of the growing oligonucleotide chain. During synthesis the DMT is removed at the "DMT-cleavage" step, and the incoming phosphoramidite, if successfully incorporated, provides the new 5'DMT end. If the phosphoramidite is not successfully incorporated, the failure sequence is acylated in subsequent synthesis steps to produce an end-capped failure oligonucleotide with no 5'DMT group. This difference provides the basis of separation with PPII columns. Since oligomers bearing a DMT-group bind strongly to reversed-phase supports, and oligomers without a DMT-group bind with less affinity, the full length sequences can be retained while the shorter failure sequences are removed with appropriate wash procedures. In overview, the PPII purification procedure proceeds as follows: First, prepare the column; Second, prepare the oligonucleotide sample; Third, load the crude oligomer solution; Fourth, remove the failure sequences; Fifth, detritylate the full-length oligomer; Sixth, exhaustively wash the support-bound oligomer; Seventh, elute the oligomer.

3.4.2.3 Sample Prep and Purification Procedures

Approximately 200 µl of the crude oligomer solution was retained for archival and comparison purposes in subsequent characterization experiments. The remaining volume of crude oligonucleotide solution, about 3 - 4 ml, was diluted with 3 volumes of distilled deionized water for a total volume of 12 – 16 ml of solution. The PPII cartridge column was prepared on a vacuum manifold by flushing the column with 10 ml of acetonitrile, followed by 10 ml of 2.0 M triethylamine acetate, pH 7.0 (Figure 9). The column was removed from the manifold, connected to a 20 ml Luer-Lok syringe and the crude oligomer solution (prepared as previously described) was manually pushed through the column at an approximate flow rate of 1 drop/sec.[70] The effluent was retained and manually passed through the column a second time. Again, the effluent was retained for the purposes of disaster recovery. The bound DMT-on oligonucleotide was washed, on a vacuum manifold, with 10 ml of 10% v/v ammonium hydroxide followed by 10 ml of distilled deionized water.[71] The oligonucleotide was then detriylated by manually pushing, with a 20 ml syringe, 10 ml of 2% v/v trifluoracetic acid in water through the column at approximately 5 ml/min. The column was then immediately rinsed with 30 ml of distilled deionized water on the vacuum manifold.[72] The detritylated oligonucleotide was manually eluted from the solid support with 2 ml of 30% v/v acetonitrile in water. The purified oligonucleotide solution was concentrated by lyophilization with a Speed-Vac lyophilizer. The resulting residue was reconstituted in 1.0 ml of deionized distilled water and a DNA concentration assay was immediately performed. The oligonucleotide was then extracted three times with 2.0 ml of diethyl ether to extract any residual trityl alcohol and then concentrated once again with a Speed-Vac lyophilizer. The purified

DNA pellet was reconstituted in water or buffer depending upon the intended purpose of the experiment and the DNA concentration measured.

3.4.2.4 Desalting, Buffer Exchange, and DNA Concentration Procedures

The PPII column was prepared as previously described. The purified oligonucleotide solution, in an unwanted buffer solution at concentration "X", was manually loaded onto the column using a 10 ml syringe at a flow rate of approximately 1 ml/min. The support bound oligonucleotide was rinsed with 30 – 40 ml of deionized water on a vacuum manifold to remove all salt ions. The oligonucleotide was manually eluted from the solid support with 2 ml of 30% v/v acetonitrile in water. The purified oligonucleotide in acetonitrile/water was concentrated by lyophilization with a Speed-Vac lyophilizer and then reconstituted in water or buffer depending upon experimental need.

3.4.2.5 PolyPakII using an Econo-Pump

Purification of the oligomer by PPII columns with an Econo-Pump is essentially the same operation as with using a vacuum manifold. However, the Econo-Pump system allows direct observation of the washing processes by monitoring the absorbance of the wash effluents at 254 nm in real-time. This visualization allows the operator to make real-time decisions with respect to wash volume and retain solutions. Use of this system should be carefully monitored to ensure that the flow-rates do not significantly exceed the manufacturer's specifications, and that the column is not allowed to run dry.

3.4.2.6 Sample Preparation and Purification Procedures

Approximately 200 µl of the crude oligomer solution was retained for archival and comparison purposes in subsequent characterization experiments. The remaining volume

of crude oligonucleotide solution, about 3 - 4 ml, was diluted with 3 volumes of distilled deionized water for a total volume of 12 – 16 ml of solution. The PPII column was prepared on a vacuum manifold by rinsing the column with 20 ml of acetonitrile, followed by 20 ml of 2.0 M triethylamine acetate, pH 7.0 (Figure 9).[73] The column was removed from the manifold, connected to a 20 ml syringe, and the crude oligomer solution (prepared as previously described) was manually pushed through the column[74] with an approximate flow rate of 1 drop/sec.[70] The effluent from this step was retained and manually pushed through the column a second time. Again, the effluent was retained for disaster recovery. The support-bound DMT-oligonucleotide was washed, using the Econo-Pump system, with 20 ml of 10% v/v ammonium hydroxide in water followed by 20 ml of distilled deionized water at a flow rate of approximately 1 – 2 ml/min (Figure 10). The DMT-oligonucleotide was detriylated by manually pushing 10 ml of 2% v/v trifluoracetic acid in water through the column at approximately 4-5 ml/min with the Econo-Detector in-line (Figure 11).[75] The column was then immediately rinsed with 30 ml of distilled deionized water with the Econo-Pump in-line. The column was dismounted from the Econo-Pump system, and the detritylated oligonucleotide was manually eluted from the solid support with 2 ml of 30% v/v acetonitrile in water. The purified oligonucleotide solution was concentrated by lyophilization with a Speed-Vac lyophilizer. The resulting residue was reconstituted in 1.0 ml of deionized distilled water and a DNA concentration assay was immediately performed. The oligonucleotide was extracted three times with 2.0 ml of diethyl ether to remove any residual trityl alcohol and then concentrated once again with a Speed-Vac lyophilizer. The purified DNA pellet

was reconstituted in water or buffer depending upon the intended purpose of the experiment and the DNA concentration measured.

3.4.3 Preparative RPHPLC

3.4.3.1 Essential Principles – HPLC on a Preparative Scale

The purification of full-length tritylated oligomers from failure sequences utilized the same principles as described for PPII columns. Oligomers bearing a DMT-group bind strongly to reversed-phase supports due to increased hydrophobic interactions, whereas oligomers without a DMT-group bind with less affinity. Under appropriate solution conditions the full length sequences can be retained on the RPHPLC column while the shorter failure sequences are removed. The purification of DNA oligomers by HPLC occurred in four distinct phases (Figure 23). First the full-length tritylated DNA was separated form the exocyclic protecting groups and N-1, N-2 etc. failure sequences. Second, the full-length DMT-oligonucleotide was detritylated using mild acidic conditions. Third, the trityl alcohol was removed by ether extraction. Fourth, the final purification of the full-length oligonucleotide was accomplished with a second round of RPHPLC (called trityl-off RPHPLC). After the final separation, the oligonucleotide was lyophilized to remove RPHPLC solvents, and then dissolved in water for dialysis against buffer or water to remove acetate salts. Besides selectivity, recovery and speed (relative to gel electrophoresis) one of the main advantages of HPLC is the ability to directly monitor the quality of the DNA separation. This enables decision-making on a rapid basis without the need for downstream analysis.

3.4.3.2 Trityl-On DNA, Prep-RPHPLC Procedures

The oligomer solution was removed from the deprotection oil bath and allowed to cool to room temperature. Occasionally the deprotection step was performed well in advance of the RPHPLC purification. Such oligomers were stored at –20 °C until further processing could occur. Oligomers stored at –20 °C, were allowed to equilibrate to room temperature prior to purification. To prevent base-mediated hydrolysis of the RPHPLC column support, all ammonium hydroxide was removed with lyophilization prior to chromatographic separation.[76] To maintain basicity during lyophilization, approximately 2 ml of triethylamine was added to the crude oligomer solution.[77] The solution was lyophilized to dryness with either a Savant Speed-Vac or a Buchi RotaVap equipped with a refrigerated vapor trap.[78] The residue was dissolved with 2.0 ml of 0.1 M triethylamine acetate pH 7.0 containing 2% acetonitrile, then filtered through a 0.2 µm syringe filter to remove any particulates. Approximately 50 µl of solution was injected onto the RPHPLC column to determine the purity of the crude oligomer and the nature of subsequent the prep-RPHPLC separation with the "TR-DNA-A" RPHPLC gradient described in the "Analytical RPHPLC" section below.[79] The remainder of the solution was loaded onto the Waters µBondapak C-18, Radial compression column and separated using the "TR-DNA-P01" gradient with the mobile phases described below (Table 6).[80] Occasionally, oligomers with high GC-content, such as CGG10 through CGG25 required a gradient with less polarity in the final phase of the HPLC gradient. These oligomers utilized the gradient "GCTR-DNA-P01" with the mobile phases described below (Table 7).[81] The major peak eluting last in the chromatogram, the DMT-on DNA peak, was collected as several fractions. Those fractions corresponding to the central portion of the DMT-on

Table 6: The "TR-DNA-P01" RPHPLC gradient.

This gradient, developed by S. Marotta, is for the primary purification of trityl-on oligomers at the 1 μmole scale.[80]

TR-DNA-P01 (for oligomers with GC content 60% or less).
All other equipment settings as specified under "Equipment Definitions."
Detector Setting: 2.0 AUFS, 280 nm
Injection Loop Size: 2.0 ml
Mobile Phase A: 0.1M triethylamine acetate pH 7.5 with 2% acetonitrile (filtered and degassed)
Mobile Phase B: acetonitrile (HPLC-grade, degassed)

TIME, min.	FLOW ml/min	%A	%B
0	4.0	90	10
0.5	4.0	90	10
15.5	4.0	75	25
25.5	4.0	75	25
27.5	4.0	90	10

Table 7: The "GCTR-DNA-P01" RPHPLC gradient.

This gradient increases the final percentage of acetonitrile to elute oligomers rich in GC sequences.

GCTR-DNA-P01 (for oligomers with high GC content)
All other equipment settings as specified under "Equipment Definitions."
Detector Setting: 2.0 AUFS, 280 nm Detector Setting: 2.0 AUFS, 280 nm
Mobile Phase A: 0.1M triethylamine acetate pH 7.5 with 2% acetonitrile (filtered and degassed)
Mobile Phase B: acetonitrile (HPLC-grade, degassed)

TIME, min.	FLOW ml/min	%A	%B
0	4.0	90	10
0.5	4.0	90	10
20.5	4.0	70	30
30.5	4.0	70	30
32.5	4.0	90	10
40	4.0	90	10

DNA peak (usually 20 – 30% acetonitrile) were retained and lyophilized to dryness. Each fraction was redissolved in 1.0 ml of deionized distilled water, filtered with a 0.45 μm syringe filter and then subsequently analyzed by either RPHPLC using the "TR-DNA-A" gradient (described in "Analytical RPHPLC") or with gel electrophoresis under denaturing conditions (Urea-PAGE). Those fractions confirmed as having the correct size either on the basis of electrophoretic migration or RPHPLC retention were pooled for subsequent detritylation. then detritylated.

3.4.3.3 Detritylation

Prior to detritylation procedures, the DNA concentration was measured with the A260 assay. A 50 μl aliquot of the pooled fractions was removed and injected onto a Waters μ-BondaPak RPHPLC column using a 50 μl injection loop with the gradient "ANAL-DNA" (described below). This chromatogram provided information about the nature of the oligomer prior to detritylation. While the chromatogram was in progress, the remainder of the fraction-pool was detritylated with the addition of an equal volume of 0.4 N acetic acid to the pool. The acetic acid acidified the pool, and hydrolyzed the acid-labile DMT group from the full-length oligomer. During this time, the acidic nature of the solution was confirmed by spotting 10 μl of solution onto a segment of pH paper. At the conclusion of the first chromatogram, approximately 15 minutes later, a second injection of 100 μl was made onto the to μ-BondaPak RPHPLC column to monitor the progress of the detritylation reaction. The second RPHPLC analysis used the same gradient. Completion of the detritylation reaction was confirmed both by the shift in retention time of the full-length oligomer and the complete disappearance of the

tritylated-DNA peak. Occasionally, a large trityl peak was also observed near the void volume of the chromatogram. Samples with incomplete detritylation were allowed to incubate at room temperature for an additional 15 minutes prior to a third RPHPLC analysis was performed. After detritylation was confirmed, the oligomer solution was "neutralized" with the addition of 2 ml of concentrated ammonium hydroxide and the basicity of the solution (pH > 8.0) confirmed by spotting 10 µl onto a segment of pH paper. The solution was lyophilized to dryness using either a RotaVap, or a Savant Speed Vac and then redissolved in deionized distilled water for subsequent ether extraction.

3.4.3.4 Ether Extraction

To one volume of oligomer solution approximately 2 volumes of diethyl ether was added and the mixture vigorously shaken for 30 seconds. The phases were allowed to separate,[82] and the ether layer was removed by aspiration with a pipette. This procedure was repeated three times, or until the aqueous layer became clear. The resulting ether-extracted oligonucleotide solution was lyophilized either with a RotaVap or Savant Speed-Vac. The residue was redissolved in 1.0 ml of water or buffer, filtered using a 0.45 µm syringe filter and analyzed by analytical RPHPLC, AXHPLC, or analytical gel electrophoresis to determine if further purification was required. Those oligomers with the proper size and purity were not processed beyond this point. Those oligomers with failure sequences, appearing either as faint bands by gel electrophoresis or low level peaks by RPHPLC, or AXHPLC, were purified a second time using the gradient "DNA-PREP-01" described below.[83]

3.4.3.5 Trityl-Off DNA, Prep-RPHPLC Procedures

Prior to chromatographic analysis an A260 assay was performed to determine the DNA concentration and yield up to this point of the purification. A 50 µl aliquot of DNA was removed from the ether-extracted, dissolved oligomer solution and injected onto the RPHPLC column, using gradient "ANAL-DNA" (described under "Analytical RPHPLC") to determine the nature of the following preparatory chromatogram.[79] The remainder of the solution was injected onto the RPHPLC column and separated with the "DNA-PREP-01" gradient using the mobile phases described below (Table 8).[80] Those oligomers with high GC contents, such as (CCG)n or (CGG)n, were purified with a modified gradient called "GCDNA-PREP-01" defined below (Table 9). The major, full-length DNA peak eluting between 10 – 15% acetonitrile for "DNA-PREP-01" gradient or 20 - 25% acetonitrile for the "GCDNA-PREP-01" gradient was collected as several fractions. Those fractions corresponding to the central portion of the DNA peak were retained and lyophilized to dryness. Each fraction was dissolved in 1.0 ml of deionized distilled water, filtered with a 0.45 µm syringe filter and then subsequently analyzed by either RPHPLC using the "ANAL-DNA" gradient (see Analytical RPHPLC) or analytical gel electrophoresis.

3.4.4 Dialysis

3.4.4.1 Principles, and Advantages

The objective of most dialysis procedures is the exchange of undesirable and relatively small ions across a semi-permeable membrane, while at the same time retention of macromolecules. To achieve this objective, it is necessary to select dialysis tubing with

Table 8: The "DNA-PREP-01" RPHPLC gradient.

This gradient, developed by S. Marotta, is for the secondary purification of DNA (trityl-off) at the 1 μmole scale.[80]

DNA-PREP (for oligomers with GC content 60% or less):
All other equipment settings as specified under "Equipment Definitions."
Detector Setting: 2.0 AUFS, 280 nm
Injection Loop Size: 2.0 ml
Column: Waters, μ-BondaPak C-18, 8 mm X 100 mm
Mobile Phase A: 0.1M triethylamine acetate pH 7.5 with 2% acetonitrile (filtered and degassed)
Mobile Phase B: acetonitrile (HPLC-grade, degassed)

TIME, min.	FLOW ml/min	%A	%B
0	4.0	97	3
0.5	4.0	97	3
24.5	4.0	85	15
34.5	4.0	85	15
39.5	4.0	97	3
50.0	4.0	97	3

Table 9: The "GCDNA-PREP-01" RPHPLC gradient.

This gradient is suitable for the purification of DNA with high GC content, due to the increase in acetonitrile near the end of the gradient.

GCDNA-PREP (for oligomers with high GC content)
All other equipment settings as specified under "Equipment Definitions."
Detector Setting: 2.0 AUFS, 280 nm Detector Setting: 2.0 AUFS, 280 nm
Column: Waters, μ-BondaPak C-18, 8 mm X 100 mm
Mobile Phase A: 0.1M triethylamine acetate pH 7.5 with 2% acetonitrile (filtered and degassed)
Mobile Phase B: acetonitrile (HPLC-grade, degassed)

TIME, min.	FLOW ml/min	%A	%B
0	4.0	97	3
0.5	4.0	97	3
30.0	4.0	75	25
34.5	4.0	75	25
39.5	4.0	97	3
50.0	4.0	97	3

pores small enough so that none of the macromolecules escape, yet large enough to allow small ions to freely diffuse into the bulk solution. Selection of the porosity begins with consideration of the size of the ions being exchanged, and the size of the macromolecule to be retained. The pore size should be, at a minimum, two to three times smaller than the macromolecule, and about 5-10 times larger than the exchanging ions.[84] As an example, the molecular mass of monobasic sodium phosphate is 120, the mass of CAG12 is 3400, and the pore size selected was 1000 MWCO. The driving force for the ion-exchange is entropy maximization, which is also described as the passive diffusion of molecules down a concentration gradient.

Dialysis has a few advantages over other techniques for buffer exchange. First, it is a true equilibrium process. While it is true that exhaustive washing can be accomplished on a PPII column, this procedure assumes rapid transfer of salt ions from the large DNA molecule to the surrounding solution. Dialysis makes no such assumptions, because the exchange of ions from high concentration to low concentration is allowed to reach equilibrium prior to solution changeover.[85] In addition, the dialysis procedure can be exhaustive – salt exchanges of one thousand- fold or even one million-fold are readily achieved with proper selection of dialysate volume and buffer tank size. The salt exchange, which describes the dilution factor for the original salt concentration, was calculated as shown below (Equation 9) where V_t is the volume of the dialysis solution ("Vtank), and V_b is the volume of the sample ("Vbag).

Equation 9: **Salt Exchange Factor = V_t/V_b**

Multiple buffer changeovers increase the overall salt exchange, and can multiply the salt exchange factors very rapidly. The salt exchange factor can be calculated as follows (Equation 10) where each term in parenthesis represents the exchange for a single buffer exchange during dialysis.[86]

Equation 10: **Salt Exchange Factor = $(V_f/V_b)(V_f/V_b)$**

The drawbacks to dialysis include both sample loss and sample dilution. Highly charged or hydrophobic molecules, such as DNA can adhere to the dialysis membranes, reducing the overall DNA purification yields.

3.4.4.2 Sample Preparation and Procedures.

Membrane selection was determined with a retentate to pore-size ratio of at least 3 or more. The smallest oligomer, the 12-mer of CAG with an approximate molecular size of 3400 Daltons was dialyzed using a cellulose acetate membrane of 1000 molecular weight cut-off (MWCO). All other oligomers, were dialyzed using the same porosity and type of membrane, thus achieving higher retentate to pore size ratios.[87] For all dialysis tubing used, a 100 cm segment was immersed in deionized distilled water, heated from 90 to 100 °C for 10 to 15 minutes and then cooled to room temperature. This pretreatment which did not degrade the cellulose acetate dialysis tubing, removed trace metal ions, destroyed any microbial growth present in the storage solution, and denatured any residual proteases/DNAses/RNAses. The dialysis tubing was then soaked in the experimentally relevant dialysis buffer for 2 exchanges over two hours. A 10 cm segment of tubing was removed and sealed approximately 2 cm from the bottom of the tube with a dialysis clip.[88] Before placement of the DNA solution into the tubing, approximately 50 µl of sample was retained for subsequent analytical purposes. The

remainder of the DNA solution, about 1 ml, was placed into the dialysis tubing and then sealed on the top end with a dialysis clip about 2 – 3 cm from the bottom clip. To ensure that no leakage occurred past the clips, both the top and bottom ends, beyond the dialysis clips, were tied with unwaxed, unflavored dental floss. The sample bag was placed into a beaker containing 1 liter of prechilled dialysis buffer at 4 °C. A magnetic stir bar was placed inside of the beaker, and the beaker set on top of a magnetic stir plate in a refrigerated "cold-box". The samples were dialyzed, with gentle stirring, for a minimum of 6 hours at 4 °C.[89] After equilibration, the buffer was discarded, and another liter of dialysis buffer was added to the beaker. The dialysis continued at 4 °C for at least another 6 hours of dialysis time.

3.4.4.3 Dialysis Yields

The overall yield of oligomer after dialysis was determined by the pre and post DNA concentrations measured by absorbance at 260 nm. The yield was calculated as shown below (Equation 11).

Equation 11: $$\%yield = ([DNA]_{postdialysis} / [DNA]_{predialysis})*(100\%)$$

3.4.5 Lyophilization

3.4.5.1 Principles

Lyophilization is a process where water and other volatile vapors are separated through evaporation and in some occasions sublimation, leaving behind nonvolatile molecules such as DNA and salts. The process depends upon a number of factors, such as the strength and volume of the vacuum and the temperature difference between the chamber temperature and the vapor trap.

3.4.5.2 Sample Preparation

Samples were checked to ensure that the solution level inside of the tube did not exceed 75% of the tube capacity to prevent the contents from spilling during lyophilization. Depending on the solution conditions the tubes were either spun in the lyophilization chamber without caps (caps-off) or with caps containing a hole in the center of the cap (caps-on). Solutions with a high probability of "bumping," such as the ether-extracted oligomers, were lyophilized "caps-on" to reduce the probability for cross-contamination. Solutions with a low probability of "bumping" were lyophilized with caps-off. Prior to starting the lyophilizer, the vapor-trap was chilled to –50 °C or colder. Lyophilized samples were reconstituted either in deionized distilled water or the buffer solution relevant to the experimental protocol.

3.5 Oligomer Characterization

3.5.1 Analytical Overview

The purity analysis of oligomers was performed with three different analytical techniques, RPHPLC, AXHPLC, or UREA-PAGE either singly or orthogonally in various combinations. The definitive assay used for the characterization of purity, size, and content was denaturing gel electrophoresis (UREA-PAGE) because of the clarity of the end-result. Unlike RPHPLC and AXHPLC, UREA-PAGE can separate N from N-1 oligomers, and was not subject to the limitations of secondary structure. In no instance were the results from a RPHPLC or AXHPLC chromatogram relied upon as the sole determinant of oligomer purity. Rather, these data were used as supplementary information to support the UREA-PAGE purity analysis results.

3.5.2 Analytical RPHPLC

3.5.2.1 Essential Principles

Reversed-phase HPLC is a separation technique that offers selectivity and good resolution in the analysis of TR-DNA, failure sequences, or full-length DNA. This separation technique was used primarily to determine the purity of in-process purification samples. One of the drawbacks of this technique is susceptibility to DNA secondary structural effects. Sequences that can form alternative secondary structures such as hairpins or quadruplexes will, depending upon solution conditions, elute as multiple peaks due to the relative hydrophobicity of each structure. This can give rise to an apparent impurity in the chromatographic profile, when in fact the impurity is simply a subpopulation of different structural forms. Furthermore, this technique is not orthogonal to the purification process and does not impart any new information about the oligomer. Any impurity that comigrates with the DNA peak during purification will also comigrate with the DNA peak during analysis because the same column, and essentially the same gradient, is used for both analysis and preparatory-scale separation. In spite of these pitfalls, RPHPLC can provide information about the presence of failure sequences, the presence of multiple structures, and the migration reproducibility of the fractionated DNA peak.

In-process purification samples were analyzed either with the "TR-DNA-A" gradient for DMT-on oligomers, or the "DNA-ANAL" gradients for DMT-off oligomers. Final, or purified DNA, samples were analyzed with either the "DNA-ANAL" or, for those prone to secondary structure and multiple peaks, the "DenDNA-ANAL" gradient in concert

with a newly-developed denaturing sample treatment protocol. The denaturing DNA sample treatment protocol is based upon the work of Arghavani et al. and avoids the secondary structural problems with other RPHPLC gradients. [90]

3.5.2.2 Sample Preparation, Procedures and Gradients

Approximately 50 µl, representing from 25 - 50 µg of DNA was lyophilized and then dissolved in 100 µl of 0.1M triethylamineacetate, pH7.0. Prior to analysis the sample was filtered through a 0.45 µm syringe filter. Approximately 50 – 100 µl of reconstituted and filtered sample was injected onto the waters µ-BondaPak RPHPLC column with the gradients and procedure-specific hardware settings specified below. Oligomers with the DMT-group attached were analyzed with the "TR-DNA-A" gradient (Table 10), whereas oligomers without the DMT-group were analyzed with the "ANAL-DNA" gradient (Table 11).[80] Samples with chromatograms indicative of secondary structural effects (Figure 24) were reanalyzed with the denaturing RPHPLC method described below (Table 12). Prior to each sample injection, 100 µl of buffer solution was injected to determine the background level and migration of peaks eluting from the column at the solution conditions determined by the gradient.

3.5.2.3 Data Analysis/Interpretation

The RPHPLC chromatograms were examined for the following properties. First, the DNA peak had to elute as a single peak with the correct retention time, as defined by the preparatory separation. Samples with chromatograms displaying failure sequences (peaks before the main DNA peak) or the wrong retention time were either discarded or repurified with RPHPLC (Figure 24). Samples with multiple peaks after the main DNA

Table 10: The "TR-DNA-A" RPHPLC gradient.

This gradient, developed by S. Marotta, allows rapid analysis of tritylated DNA.[80]

TR-DNA-A gradient:
All other equipment settings as specified under "Equipment Definitions."
Detector Setting: 0.5 AUFS, 260 nm
Injection Loop Size: 50 µl or 100µl
Column: Waters, µ-BondaPak C-18, 8 mm X 100 mm
Mobile Phase A: 0.1M triethylamine acetate pH 7.5 with 2% acetonitrile (filtered and degassed)
Mobile Phase B: acetonitrile (HPLC-grade, degassed)

TIME, min.	FLOW ml/min	%A	%B
0	4.0	90	10
0.5	4.0	90	10
5.5	4.0	70	30
6.0	4.0	70	30
8.0	4.0	90	10
15.0	4.0	90	10

Table 11: The "ANAL-DNA" RPHPLC gradient.

This gradient developed by S. Marotta allows efficient analysis of detritylated DNA.This gradient has a shallow slope and enables separation of those peaks, such as detritylated DNA, that tend to be poorly resolved with the "TR-DNA-A" gradient.[80]

ANAL-DNA gradient
All other equipment settings as specified under "Equipment Definitions."
Detector Setting: 0.5 AUFS, 260 nm
Injection Loop Size: 50 µl or 100µl
Column: Waters, µ-BondaPak C-18, 8 mm X 100 mm
Mobile Phase A: 0.1M triethylamine acetate pH 7.5 with 2% acetonitrile (filtered and degassed)
Mobile Phase B: acetonitrile (HPLC-grade, degassed)

TIME, min.	FLOW ml/min	%A	%B
0	4.0	97	3
0.5	4.0	97	3
5.5	4.0	83	17
6.0	4.0	83	17
7.0	4.0	97	3
15.0	4.0	97	3

Table 12: The "DenANAL-DNA" RPHPLC gradient.

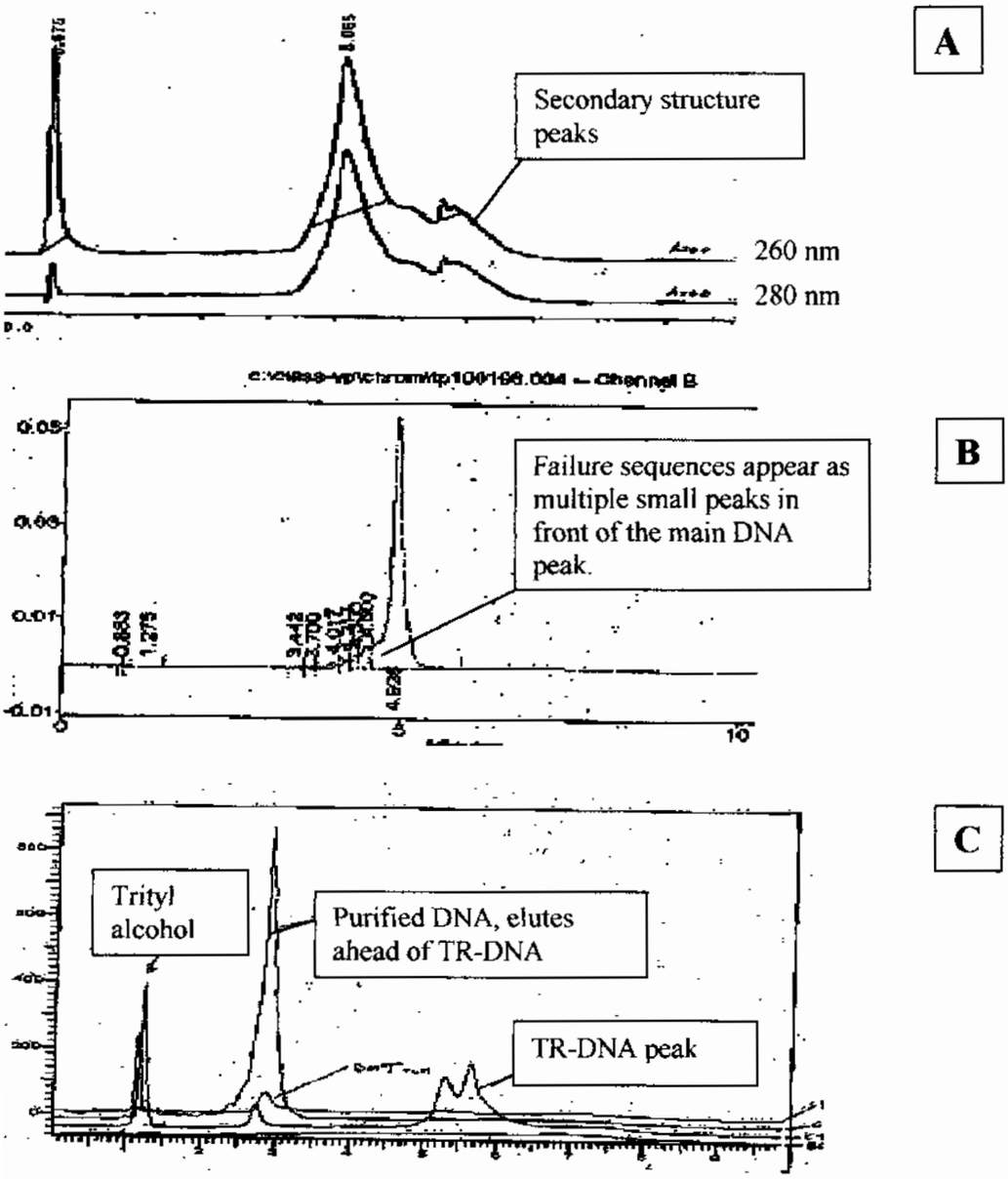
DenANAL-DNA" RPHPLC
Sample treatment specific to this method:
Approximately 50 µg of DNA was lyophilized, to concentrate the DNA, then reconstituted with 100 µl of 33% formamide in water. The samples were heated at 90 – 100 °C for 10 minutes on a heat block, then rapidly cooled (less than 1 minute) to 4 °C using an ice bath. The samples were injected onto the RPHPLC column (at room temperature) with the gradient below.

All other equipment settings as specified under "Equipment Definitions."
Detector Setting: 0.1 AUFS, 260 nm
Injection Loop Size: 50 µl or 100µl
Column: Waters, µ-BondaPak C-18, 8 mm X 100 mm
Mobile Phase A: 0.1M triethylamine acetate pH 7.5 with 2% acetonitrile (filtered and degassed)
Mobile Phase B: acetonitrile (HPLC-grade, degassed)

TIME, min.	FLOW ml/min	%A	%B
0	4.0	97	3
0.5	4.0	97	3
5.5	4.0	80	20
6.0	4.0	80	20
7.0	4.0	97	3
12.0	4.0	97	3

Figure 24: Typical RPHPLC chromatograms.

A. A sample of DNA with significant secondary structure has the appearance of an impure sample containing multiple peaks eluting after the main peak. **B.** A sample of DNA with a high level of failure sequences. **C.** An overlay of a typical TR-DNA chromatogram with a purified DNA chromotogram. The RPHPLC chromatograms were examined for the following properties: the correct retention time, absence of failure sequences, peak area relative to the injection amount, and lack of multiple peaks after the main DNA peak.



peak (indicating secondary structures) were analyzed a second time with the denaturing DNA protocols (Figure 24). When analyzed under denaturing conditions, the same rules were applied – sample with failure sequences were either discarded or repurified depending upon the anticipated yield. Samples with "stubborn" secondary structure were analyzed by denaturing gel electrophoresis or in rare cases AXHPLC.

3.5.3 Analytical Anion Exchange HPLC (AXHPLC)

3.5.3.1 Essential Principles- Why Use AXHPLC

Anion exchange HPLC separates samples on the basis of overall negative charge. The stationary phase contains a positively charged tertiary amine moiety that can bind negatively charged anions, and polyanions such as DNA, with high affinity. In practice the anion is bound to the solid support, then eluted away as the concentration of competing anion, such as sodium chloride, is increased. Anion exchange HPLC offers better resolution and selectivity with respect to DNA oligomers than does RPHPLC and is capable of fully resolving "n" from "n-1" failure sequences. Like RPHPLC however, this technique is limited to samples without secondary structural effects. In addition, the capacity of most commercially available analytical AXHPLC columns limits their application solely to the analysis, as opposed to the preparation, of DNA oligomers.[91] This technique was used occasionally to provide orthogonal information about the purity of DNA solutions.

3.5.3.2 Sample Prep

Approximately 25 - 50 μg of DNA was lyophilized to remove any volatile buffer components, then dissolved in 100 μl of 10 mM NaH_2PO_4. The DNA sample was

filtered with a 0.45 µm syringe filter, and injected onto the AXHPLC column with the gradient "AX-ANAL-DNA" specified below (Table 13). Prior to sample injection, a blank sample of buffer alone was injected onto the column to determine the background level of extraneous peaks.

Table 13: The "AX-ANAL-DNA" AXHPLC gradient.

All other equipment settings as specified under "Equipment Definitions."
Detector Setting: 0.5 AUFS, 260 nm
Injection Loop Size: 50 µl or 100µl
Column: Hychrom SynchroPak AX300, 4.6 mm by 250 mm
Mobile Phase A: 25 mM Tris/HCl pH 8.0 10% CH3CN (filtered and degassed)
Mobile Phase B: 25 mM Tris/HCl, 1000 mM Nacl pH 8.0 10% CH3CN (HPLC-grade, degassed)

TIME, min.	FLOW ml/min	%A	%B
0.0	1.5	100	0
1.0	1.5	100	0
20.0	1.5	0	100
23.0	1.5	0	100
25.0	1.5	100	0
30.0	1.5	100	0

3.5.3.3 Data Analysis/Interpretation

All data was analyzed with the same criteria applied for RPHPLC. Namely, all chromatograms were inspected for the correct retention time and the presence of failure sequence peaks. Any samples containing failure sequence peaks were either discarded or repurified by RPHPLC depending upon the anticipated yield. Those samples with the correct retention time and purity were retained.

3.5.4 DNA Concentration Determination - The A260 Assay

3.5.4.1 Essential Principles

Spectrophotometry is one of the most used, and perhaps abused, of bioanalytical techniques.[92] It is a nondestructive technique, and affords some degree of specificity that is dependent upon the characteristic spectrum of each absorbing species. If large differences exist between the absorption spectra of two different species, then the spectrophotometric behavior of one component can be effectively isolated and observed. If this behavior can be correlated with changes to the molecule, such as denaturation in the case of DNA, then there exists an opportunity for direct measurement of phenomena occurring at the molecular level.

According to the electromagnetic theory of photon radiation, a light wave consists of electric and magnetic fields that are in-phase but perpendicular to each other and to the direction of propagation (Figure 28). The length of the wave, λ, is related to the velocity of light in a vacuum, c, and the frequency, ν (Equation 12).

Equation 12: $\lambda = c/\nu$

Light interacts with molecules primarily through the oscillating electric field, and can be thought of as a quantized energy source. The amount of energy in a photon is equivalent to the product of Planck's constant, h, and the frequency (Equation 13).

Equation 13: $E = h\nu$

Absorption of light energy occurs when the energy of the photon corresponds to a quantized energy difference between a ground state and excited state of a molecular

orbital electron (Figure 25). While the absorption energy levels derive primarily from transitions between energy levels of electrons, transitions between different vibrational, rotational, and translational energy states can also occur. A large number of biological macromolecules, such as DNA, absorb light within the visible to ultraviolet regions of the electromagnetic spectrum due primarily to the amount energy required to induce a π to π^* molecular orbital transition. Nucleic acids absorb UV light as the π-electrons of the aromatic rings undergo transitions towards higher energy states. The energy that can be absorbed by the ring system is quantized - each ring system will absorb energies corresponding to a distinct wavelength (Figure 25).

The absorption of light is related primarily through the Beer-Lambert/Bouguer law. According to Bouguer, and later by Lambert, the intensity of transmitted light is related to the intensity of the incident light according to the following relationship (Equation 14),

Equation 14: $\log(I/I_o) = -\alpha b$ **or** $I/I_o = e^{-\alpha b}$

where I is the intensity of the transmitted light, I_o is the intensity of the incident light, α is the absorptivity of the medium, and x is the thickness of the medium.[93] According to Beer the absorptivity is equal to the molar absorptivity ε times the molar concentration c, thus (Equation 15)

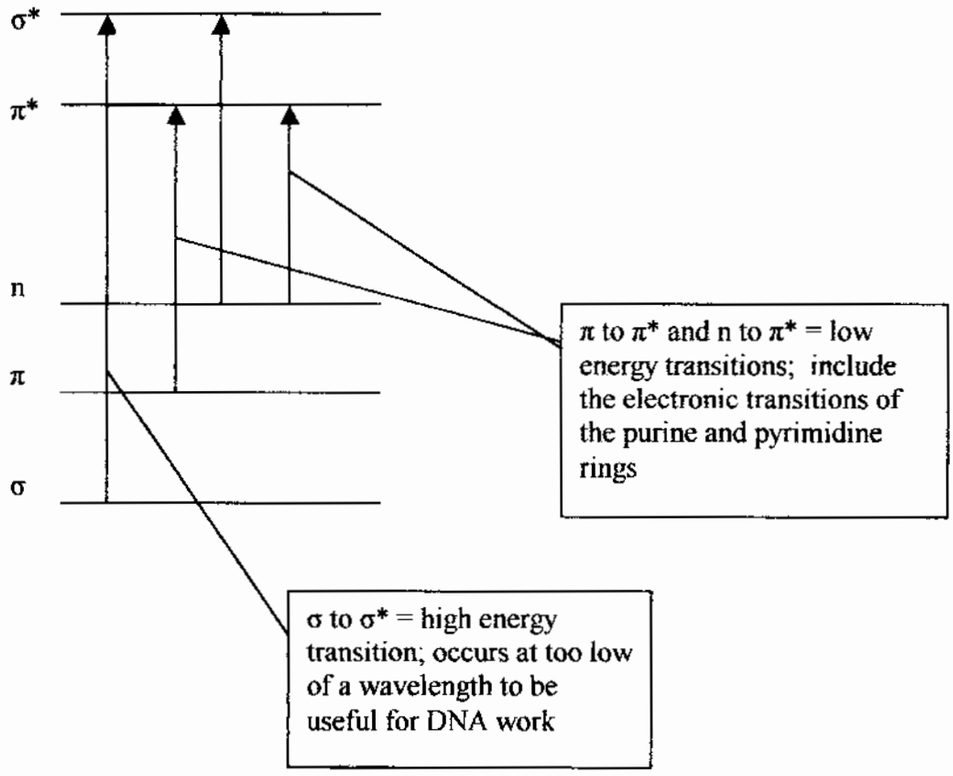
Equation 15: $\alpha b = \varepsilon c b$

where ε is the molar absorptivity (units l/(mol cm), c is the concentration and b is the path length then the Bouguer/Lambert relationship can be redefined with (Equation 16)

Equation 16: $\log(I/I_o) = \varepsilon c b$

Figure 25: Absorption of electromagnetic radiation.

Absorption of an electromagnetic wave by an electron occurs at discrete, quantized energy levels that are related to the energy of the photon. The energy absorbed excites an electron from the ground state to the next highest, or in some cases higher, energy levels.



The negative logarithm of the transmittance term is defined as the absorbance giving as the final form, Beer's law (Equation 17).

Equation 17: $A = \varepsilon c b$

3.5.4.2 The A260 Assay vs. Spectral Scan Assay

Spectral scan assays, defined as a spectrophotometric scan from 400 nm to 200 nm were used whenever information regarding the wavelength of maximum absorbance, the DNA concentration (derived from absorbance at 260 nm), purity (ratios of absorbance at 260 nm and 280 nm), and the overall quality of the DNA (single or multiple peaks) was required. Thus, from a single scan, a lot of information regarding the purity, the condition, and the concentration of the oligomer could be obtained. However, single wavelength measurements, performed typically at 260 nm, offered greater through-put and were used whenever noncritical DNA concentration measurements were required.

3.5.4.3 Extinction Coefficients

The extinction coefficient in Beer's law relates the absorptivity of a solution to the concentration (assuming a constant path length of 1 cm). Because extinction coefficients are related to the absorbance of the DNA, which in turn is related to the structure of the DNA molecule, any changes to the solution environment that change DNA structure (pH, temperature, counterion concentration) will change the extinction coefficients, assuming no changes to the DNA concentration.

Extinction coefficients were determined using one of three methods depending upon the required accuracy. In the first method, the extinction coefficient was determined

empirically (referred to as an Empirical Extinction Coefficient) based upon experimental measurements. The amount of DNA corresponding to approximately 1.0 absorbance unit at 260 nm was scanned with a spectrophotometer from 400 nm to 200 nm, and the absorbance at 260 nm was noted. The sample was then completely digested to free nucleotides with snake venom phosphodiesterase. The absorbance of the free nucleotide solution at 260 nm was measured from a spectrophotometric scan between 400 and 200 nm. The original DNA concentration was calculated from the free nucleotide concentration using the known molar extinction coefficients for the free nucleotides, and the molar ratios of each nucleotide (specified by the sequence). With the original DNA concentration known in terms of free nucleotides, the extinction coefficient for the native DNA solution was derived using the original (undigested) DNA absorbance and Beer's law. This method represented a model-free approach because the extinction coefficient was determined directly from DNA in the native state and in the solution environment specific to the experiment under consideration. However, this method was subject to the accuracy of the DNA absorption measurements, was tedious, and was relevant only to the solution conditions of the original DNA measurement. A second and more general method calculated the extinction coefficient on a theoretical basis from the sum of individual molar absorptivities (Equation 18) for all nucleotides in the sequence (Summed Extinction Coefficient or SEC,).[94] For each equation listed, C is DNA concentration in pmol/μl, N_X where X = A, C, G or T, is the number of nucleotides present in the sequence under consideration, A_{260} is the absorption at 260 nm, and ε is the molar extinction coefficient.

Equation 18: $\varepsilon = (15{,}200N_A) + (7050N_C) + (12010N_G) + (8400N_T)$

$$C\ (pmol/\mu l) = A260\ X\ \{100/[(1.5N_A) + (0.71N_C) + (1.2N_G) + (0.84N_T)]\}$$

This approach assumed that all nucleotides in a polynucleotide structure have the same or similar molar absorptivities as the free nucleotides in solution. This approximation is correct if the DNA does not form secondary structures. The accuracy of this method decreased with secondary structure formation because base stacking interactions between adjacent bases decreases the absorptivity of the nucleotides. In spite of this drawback the summed extinction coefficient was used to calculate the DNA concentration for most oligonucleotides due to the relative accuracy and universality of this method. The third method utilized an averaged extinction coefficient, (AEC) based on the averaged absorptivities of the free nucleotides, of 27 μg/ml of DNA per absorbance unit at 260 nm for single stranded DNA, and 20 μg/ml of DNA per absorbance at 260 for double-stranded DNA (Equation 19:).[94]

Equation 19: $C\ (\mu g/ml) = A_{260}/0.027$ **(single stranded DNA)**

This method, while the least accurate of all three methods, provided convenience when quick and routine calculations of DNA concentration were required. In practice, the extinction coefficient used to determine concentration depended on the required accuracy of the measurement and the prevailing experimental objectives. For experiments where the DNA concentration was critical the summed molar extinction coefficient specific to each sequence was utilized.[95] For assays where the DNA concentration was not as critical, such as gel electrophoresis or crude yield determination, the averaged extinction coefficient was used.[96]

3.5.4.4 Sample Preparation and Spectral Scan Procedures

Prior to assay, all pipettors were checked for accuracy by comparing the mass of the dispensed volume of water, averaged across five replicates, against the marked volume on the pipettor. After correcting for the density of water at the ambient temperature, those pipettors with errors greater than 2% were recalibrated, using either the manufacturers procedure or by simply adjusting the dispensed volume setting until the correct water-mass of water was obtained.[97] To eliminate possible sources of error arising from mismatched cuvettes the buffer solution was scanned from 400 nm to 200 nm prior to sample addition in each measurement cuvette. The reference for all scans was air.[98] Typically, 490 µl of 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1mM EDTA, pH 7.0, 100 mM NaCl was placed inside of the cuvette and scanned from 400 nm to 200 nm in 2 nm steps. Because some samples of DNA may have precipitated due to the storage conditions,[99] all samples were heated to 90 – 100 °C for 5 –10 minutes to resolvate precipitated DNA. The samples were mixed using a vortex mixer[100] then rapidly cooled to room temperature by centrifugation in a microcentrifuge for 5 minutes. Centrifugation not only rapidly cooled the DNA samples, but also pelleted any suspended particulates to the bottom of the microcentrifuge tube.[101] Approximately 10 µl of DNA solution was removed[102] and added to the cuvette containing 490 µl of phosphate buffer that was previously scanned (total sample dilution was 1:50).[103] The absorbance of the DNA sample was measured at 25 °C with air in the reference side from 400 nm to 200 nm in 1 nm increments, with either 1.0 nm bandwidth (for 2 mm wide cells) or 2.0 nm bandwidth (for 4 mm wide cells)[104], and the data were overlaid on top of the blank buffer scan. The maximum wavelength, the absorbance at 260 nm and the absorbance at 280 nm were

recorded for the sample and blank spectra. The overall quality of the DNA sample as reflected in the spectral scan was also noted. Such parameters included the overall shape of the spectrum,[105] the number of peaks,[106] and the noise level near 400 nm (Figure 26).[107] The relative purity of the DNA was calculated from the ratio of the absorbance at 260 nm divided by the absorbance at 280 nm.[108] The concentration of DNA was calculated from the blank-corrected absorbance value at 260 nm, multiplied by the dilution factors (Equation 5) and divided by the extinction coefficient determined by either the SEC or AEC methods (Equation 20).

Equation 20: $$[DNA] = (A_{260})\,(\text{dilution factor}) / \varepsilon$$

Those samples with absorbance values below 0.1 or above 1.5 at 260 nm were repeated with lower or higher dilutions, respectively, than described.

3.5.4.5 Sample Preparation, Single Wavelength Measurement Procedures

This measurement method offered speed and convenience at the expense of information, and was used mainly for non-critical determinations of DNA concentration. To eliminate possible sources of error arising from mismatched cuvettes the absorbance of the buffer solution (typically, 490 µl of 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1mM EDTA, pH 7.0, 100 mM NaCl) was measured at 260 nm vs. air prior to sample addition in each measurement cuvette.[98] Because some samples of DNA may have precipitated due to the storage conditions,[99] all samples were heated to 90 – 100 °C for 5 –10 minutes to resolvate any precipitated DNA. The samples were mixed using a vortex mixer[100] then rapidly cooled to room temperature by centrifugation in a microcentrifuge for 5 minutes. Centrifugation not only rapidly cooled the DNA samples, but also pelleted any suspended

Figure 26: A spectral scan of a DNA oligomer.

Spectral scans for all oligomers are examined for the overall nature of the spectra, the wavelength of maximum absorbance, the absorbance at 260 nm, absorbance at 280 nm, and the number of peaks. Shown below is an overlay of the spectral scan for pre-DSC (red) and post-DSC (blue) samples of the oligomer CAG10 . The spectra for both samples in this example are nearly super-imposable, suggesting no change in DNA concentration.

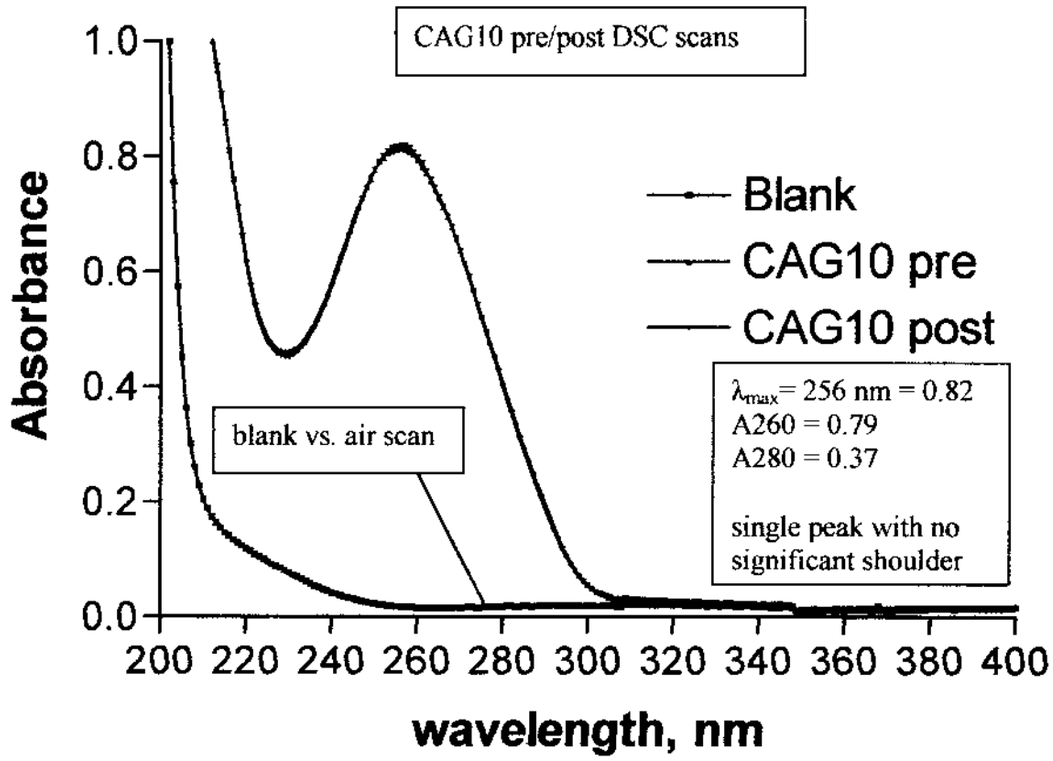
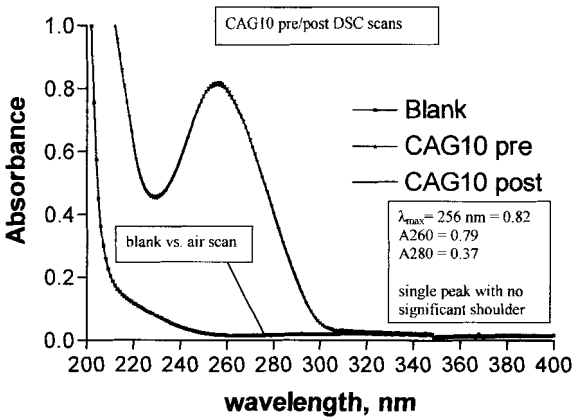


Figure 26: A spectral scan of a DNA oligomer.

Spectral scans for all oligomers are examined for the overall nature of the spectra, the wavelength of maximum absorbance, the absorbance at 260 nm, absorbance at 280 nm, and the number of peaks. Shown below is an overlay of the spectral scan for pre-DSC (red) and post-DSC (blue) samples of the oligomer CAG10 . The spectra for both samples in this example are nearly super-imposable, suggesting no change in DNA concentration.



particulates to the bottom of the microcentrifuge tube.[101] Approximately 10 µl of DNA solution was removed[102] and added to the cuvette containing 490 µl of phosphate buffer previously measured.[103] The total dilution was 1:50. The absorbance of the DNA sample was measured at 260 nm and the DNA concentration calculated as described previously. Those samples with absorbance values below 0.1 or above 1.5 at 260 nm were repeated with lower or higher dilutions, respectively, than described.

3.5.5 Analytical Gel Electrophoresis: Urea-PAGE

3.5.5.1 Essential Principles Denaturing Gel Electrophoresis

Throughout this project, denaturing gel electrophoresis was the definitive analytical tool used for purity and size analysis. Because the samples are denatured prior to analysis and maintained in a denatured state through separation, all secondary structural effects are eliminated. Samples that may have eluted as multiple peaks in a RPHPLC chromatogram, suggestive of secondary structure, were often found to be composed of a single band by gel electrophoresis. Secondly, the ability of this technique to completely resolve full-length from N-1 failure sequences, the superior sensitivity, the low sample loads, and the rapidity of sample analysis made this technique the definitive analytical technique in the analysis of DNA samples.

Electrophoresis is the process through which ions, such as DNA, are separated in an electric field due to their different mobilities. Factors that affect ion mobility are the charge of the molecule (q), the voltage gradient of the electric field (E), and the frictional coefficient of the ion as reflected in the relative mass or size (f)[109]. Mobility of an ion or

particulates to the bottom of the microcentrifuge tube.[101] Approximately 10 µl of DNA solution was removed[102] and added to the cuvette containing 490 µl of phosphate buffer previously measured.[103] The total dilution was 1:50. The absorbance of the DNA sample was measured at 260 nm and the DNA concentration calculated as described previously. Those samples with absorbance values below 0.1 or above 1.5 at 260 nm were repeated with lower or higher dilutions, respectively, than described.

3.5.5 Analytical Gel Electrophoresis: Urea-PAGE

3.5.5.1 Essential Principles Denaturing Gel Electrophoresis

Throughout this project, denaturing gel electrophoresis was the definitive analytical tool used for purity and size analysis. Because the samples are denatured prior to analysis and maintained in a denatured state through separation, all secondary structural effects are eliminated. Samples that may have eluted as multiple peaks in a RPHPLC chromatogram, suggestive of secondary structure, were often found to be composed of a single band by gel electrophoresis. Secondly, the ability of this technique to completely resolve full-length from N-1 failure sequences, the superior sensitivity, the low sample loads, and the rapidity of sample analysis made this technique the definitive analytical technique in the analysis of DNA samples.

Electrophoresis is the process through which ions, such as DNA, are separated in an electric field due to their different mobilities. Factors that affect ion mobility are the charge of the molecule (q), the voltage gradient of the electric field (E), and the frictional coefficient of the ion as reflected in the relative mass or size (f)[109]. Mobility of an ion or

macromolecule (R_f) in an electrophoretic gel is defined as the distance migrated from the top of the gel relative to a defined marker such as a molecular weight marker or dye front.

Equation 21: $\mathbf{R_f = qE / f}$

Fractionation of different molecules occurs because of the sieving effect of the gel medium and the charge to mass ratio of each species being separated. The sieving effect is such that large molecules become entrapped within the matrix of the gel, and are retarded relative to smaller ions. Denaturing electrophoresis of DNA differs slightly from denaturing electrophoresis used for proteins. In denaturing electrophoresis for DNA, the denaturant is formamide or urea. Both compounds denature DNA by competing for hydrogen-bond interactions that serve to stabilize the DNA secondary structure. Thus, under denaturing conditions, each DNA molecule has a similar size (stokes radius) to mass ratio. This differs from DNA that can form secondary structures, such as hairpins versus the single strand form of the same sequence. Though both forms have the same mass, the former has a smaller Stokes radii, and less frictional resistance relative the latter from. Because most of the charge in a DNA molecule comes from the phosphate backbone, the denaturing gel electrophoresis also imparts a constant charge to size ratio for all DNA molecules under consideration. Thus DNA of several different sizes will migrate on the basis of size, and directly give information about the length of the DNA under analysis. Denaturing gel electrophoresis for proteins differs from that used for DNA with respect to the denaturant (sodium dodecyl sulfate) and the mechanism through which denaturation is achieved. However, the separation mechanism (i.e., constant charge/mass ratio; separation on size) remains the same.

The sieving effect of gel electrophoresis was related to the pore size of the gel. The pore size was related to the percent acrylamide concentration, %C, and percent crosslinking, %T, of the electrophoretic gel. The percent cross-linking, defined by the acrylamide:bis-acrylamide (cross-linker) ratio, was held constant at 19:1 acrylamide:bisacrylamide.[110] The pore size, and therefore the sieving effect, was manipulated through the acrylamide concentration. Depending upon experimental the objective most oligomers were analyzed with a 20% polyacrylamide gel.[111] A guide to the relative size of DNA that can be separated as a function of polyacrylamide concentration is shown (Table 14).[112]

Table 14: A guide to the optimal acrylamide concentration for PAGE.

A guide to the optimal DNA size ranges that can be analyzed as a function of percent polyacrylamide concentration is shown below. The Urea-PAGE gels are defined as 8M Urea in buffer relevant to experiment with the polyacrylamide concentration listed. Native Gels buffers will vary, and only the percent acrylamide is given. Note that the upper limits of the size ranges specified are optimistic. Good resolution (as opposed to visualization of a band) between the N and N-1 failure of these higher molecular weight species will require polyacrylamide gels of lower percentages than those listed.

% Polyacrylamide	DNA Size Range, UREA-PAGE	DNA Size Range, Native
5	60-400	90-2000
10	30-300	40-1600
15	20-250	--
20	5-200	18-1600

Unless the DNA (or protein) of interest was colored, the end result of any electrophoretic separation must be visualized by some technique. For all experiments, DNA visualization was achieved with the DNA-specific fluorescent staining dye, Syber-Green I (Molecular Probes Inc. Eugene Oregon).[113] This technique, like most others used for analysis, was subject to the limitations of linearity and limit of detection. For all

measurement techniques there exists a threshold, below which no meaningful measurements can be made above the level of background noise. This is the limit of detection. Similarly, there also exists a linear range of values through which the response of the detection system obeys a linear relationship to the apparent dose of sample. A common mistake made by many practitioners of electrophoresis is to load a sample at a single concentration for purity analysis without any regard to the LOD or LR parameters. This is incorrect for two reasons. With any crude oligomeric DNA sample, there will be a large population of full-length oligonucleotides, and a smaller sub-population of failure sequences. If the DNA sample is loaded, such that the amount of the full-length population is near the limit of detection, the failure sequences, being of much lower percentage will be beyond the LOD will not be detected, giving rise to a higher apparent purity. Conversely, if the DNA sample is loaded near the top of the linear range, the full-length oligomer band will not stain to the same extent as the failure sequence bands that are within the linear range giving an apparently lower purity. For this reason, samples for purity analysis were analyzed as a titration range of two or three different concentrations. Crude purity estimates were made from those samples which were in between the upper and lower limits of detection.

3.5.5.2 Sample Preparation and Experimental Procedures for Urea-PAGE

Volumes corresponding to approximately 3, 0.3, 0.03 and 0.003 +/- 10% µg of DNA were placed into the bottom of 500 µl microcentrifuge tubes and lyophilized to dryness.[114] At the same time, 1 µl of molecular size markers (G/BRL 10 bp ladder) representing approximately 1 µg of DNA was dispensed into the bottom of two 500 µl

microcentrifuge tubes. During lyophilization, a 20% acrylamide (19:1 acrylamide:bisacrylamide) gel solution containing 8.0 M urea, 40 mM Tris-acetate, 1 mM EDTA, pH 8.0 (1X TAE) was prepared. To 25 ml of gel solution 100 µl of 10 % ammonium persulfate, and 50 µl of TEMED was added and immediately mixed. The gel solution was immediately poured into a Novex II casting stand of dimensions 1 mm x 80 mm x 80 mm. After insertion of the 10-well comb the gel was set aside to polymerize undisturbed. During polymerization, the water bath used to maintain the buffer tank temperature at 60 °C was set to 75-80 °C and the electrophoresis buffer, 40 mM Tris-acetate, 1 mM EDTA, pH 8.0 (1X TAE) was preheated to 60 °C on a magnetic stir plate equipped for heating. Upon complete lyophilization, the samples were reconstituted with 10 µl of 1X TAE, 90% formamide, 0.1% bromophenol blue, 0.1% xylene cyanol (SHU Denaturing Gel Sample Buffer) then heated at 90 – 100 °C for 10 minutes. During the sample heating time, the 60 °C TAE pH 8.0 buffer was poured into the electrophoresis tank, a stir bar was inserted inside of the tank, the external water bath was connected to the heating/cooling "coils" inside the tank (Figure 12), and the entire tank was set on top of a magnetic stir plate to thermally equilibrate for 10-15 minutes. Upon equilibration, the samples were loaded, hot,[115] onto the gel such that the molecular size markers occupied the outermost (lanes 1 and 10) lanes. Denaturing sample buffer solution was loaded into any lane not containing a DNA sample.[116] The samples were electrophoresed at 50 - 60 °C, 80 –125 V constant voltage, 1X TAE pH8.0, with gentle stirring, for approximately 1.0 – 1.5 hours or until the lower dye front was within 0.5 cm of the bottom of the gel cassette. Upon completion, the gel was removed from the cassette, and immediately immersed in a plastic staining tray[117] containing 50 ml of 1X TAE pH 8.0

with 5 µl of Syber-Green I staining dye. The tray was transferred to a rotating shaker, covered with foil then gently rotated at room temperature for 15 minutes.[118] The stained DNA bands were visualized with a UV-transilluminator and all photography accomplished with an AlphaImager digital camera system.

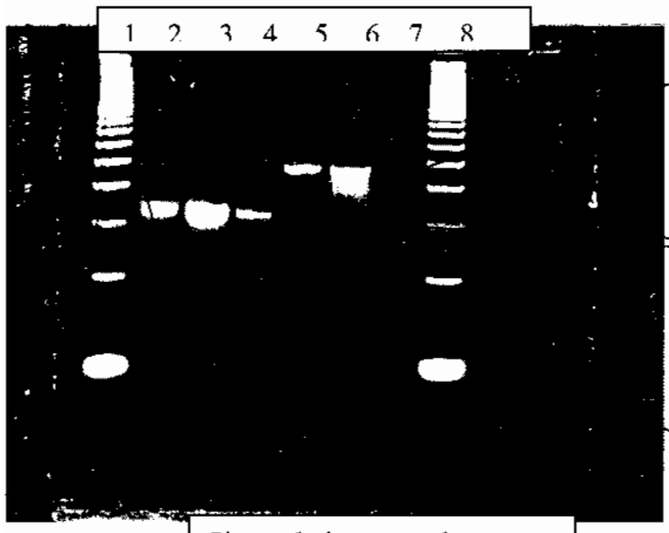
Procedures for analysis of oligonucleotides with the Hoeffer gel electrophoresis system differed little from those used for the Novex Excell II electrophoresis system. The acrylamide volume for the larger gel system was adjusted to 50 ml (approximately 40 ml required for the larger gel). The efficiency of the heating/cooling core was such that preheating of the electrophoresis buffer on a heating plate was not necessary. The buffer was thermally equilibrated to 60 °C in the Hoeffer electrophoresis tank using an external water bath as the heat source and and the Hoeffer electrophoresis heating/cooling core as the heat exchanger. In addition, the volume of the Syber-Green I staining solution was adjusted to accommodate the larger gel size. All other procedures remained the same as previously described.

3.5.5.3 Urea-PAGE Data Analysis

All gels were inspected for satisfactory resolution and detection as shown by the molecular weight markers. Gels producing unusual migration patterns ("smiles or frowns) faint bands, or poor resolution were repeated.[119] All samples were inspected for the correct oligomer size relative to the molecular weight markers, the absence of failure sequences below the main oligomer band (for those samples loaded within the LOD and linear range), relative band intensity, and any other unusual electrophoretic patterns (Figure 27). Several types of information were sought from this analysis. First, the

Figure 27: Denaturing UREA-polyacrylamide gel stained with Syber-Green I.

All samples separated on a denaturing polyacrylamide gels are inspected for purity, as indicated by the absence of failure sequences within the linear range of the assay, overall size, as indicated by the relative migration with respect to the molecular size markers, detection response, as indicated by the relative staining intensity of the DNA band, and linear range, as shown by the overall response of a titrated sample shown below. **A.** A denaturing polyacrylamide gel containing several high purity samples. Lanes 1, 8 molecular weight markers; Lane 2, Crude 30-mer prior to purification; Lanes 3,4, purified 30-mer; Lane 5, crude 45-mer prior to purification, lanes 6, 7 purified 45-mer. **B**. A denaturing polyacrylamide gel containing several low-purity samples after purification.

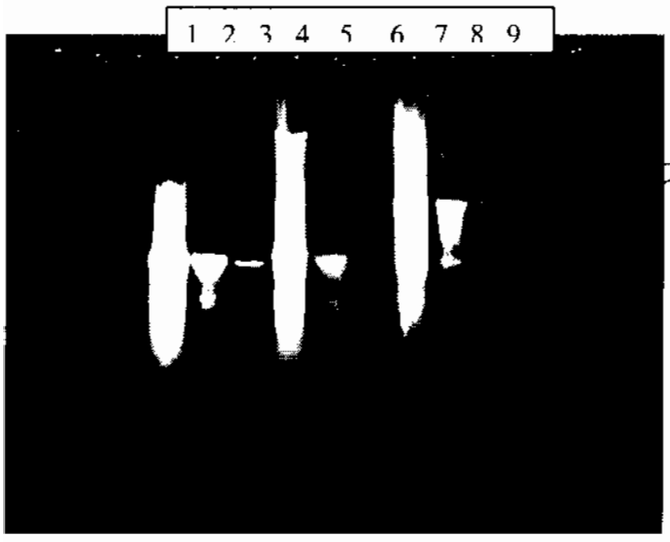


Linear range for purity analysis – last sample of titration range is too low.

Relative staining intensity between oligomers with the same sample load amount. Comparisons are best drawn between same-sized oligomers.

No bands corresponding to failure sequences

Size, relative to markers



Bands above/bands below = "junk-mer". Note this lane is overloaded and cannot be used for purity analysis. The next lane, within the linear range, shows the same banding pattern – REJECT OLIGO.

correct size of the oligomer was verified relative to the migration patterns of known DNA molecular weight markers. Second, the purity of the oligomer was verified by the absence of failure sequence bands migrating below the main oligomer band. Third, the staining intensity indicated the relative level of depurination. Depurination results in the loss of purine bases, such as guanine, from the oligonucleotide chain due to hydrolysis. Oligomers with extensive depurination produced low intensity bands relative to isolength oligomers at similar sample loads. Slight differences in staining intensity were expected, however, large differences in staining intensity were sufficient reason for rejection. Fourth, and finally, the absence of high molecular weight aggregates or unusual structures was verified. Samples with the correct size, band intensity and without failure sequences or unusual structures were retained.

3.6 Structural Characterization

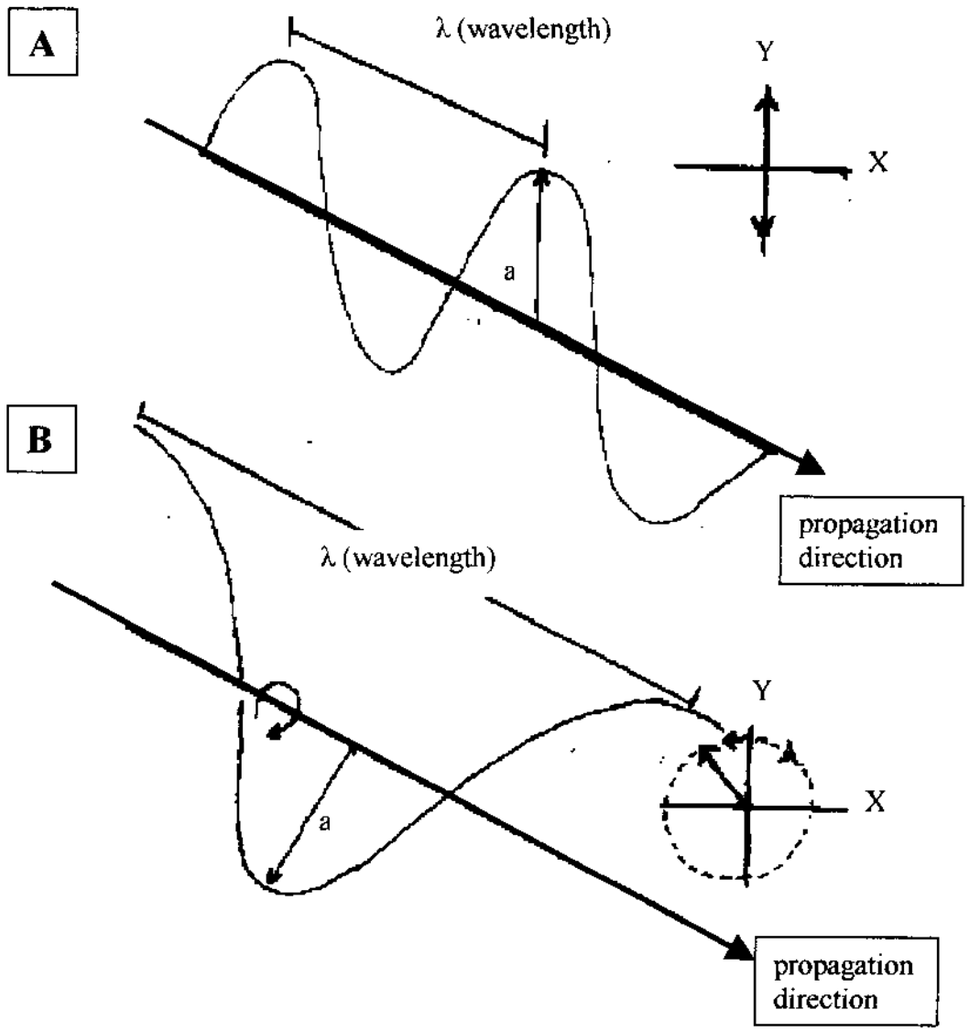
3.6.1 Circular Dichroism Spectra

3.6.1.1 Essential Principles

The electromagnetic spectrum consists of waves composed of electric and magnetic fields. The component field vectors are perpendicular to each other as well as the direction of propagation, and have an associated frequency or wavelength characteristic of the energy of the radiation under consideration (Figure 28). With plane-polarized light, the direction or plane of oscillation of the electric and magnetic fields are fixed about a selected plane of reference. Circularly polarized light differs from plane-polarized light with respect to the plane of oscillations. The amplitude of the magnetic

Figure 28: A comparison between plane and circularly polarized light.

A. The amplitude of the electric and magnetic field vectors for plane polarized light oscillates vertically about the axis of propagation (only one field is shown). Inset: The vertical oscillations are shown as the up/down arrow as with the axis of propagation perpendicular to the page.
B. The amplitude of the electric and magnetic field vectors for circularly polarized light remains fixed and rotates around the axis of propagation (only one axis shown). Inset: The amplitude rotates in a left or right circular fashion around the axis of propagation, that is perpendicular to the page.



and electric fields does not vertically oscillate about the axis of propagation, but rather rotates around the axis of propagation in either a clockwise (right-handed) or counterclockwise (left handed) direction (Figure 28). If two circularly polarized light beams are combined in equal proportions, then the net result will be plane-polarized light. However, if two unequal proportions of light are mixed, such as when passing through an optically active molecule, the result will be elliptically polarized light characterized by both an ellipticity and direction about the main axis. All samples that posses a chromophore with optical activity will absorb left-handed and right-handed circularly polarized light differently. Because absorption spectroscopy obeys Beer's law, circular dichroism ellipticity measurements also possess the same relationship as shown below (Equation 22),

Equation 22: $$A_L(\lambda) - A_R(\lambda) = [\varepsilon_L(\lambda) - \varepsilon_R(\lambda)]\, bc$$

where $A_L(\lambda)$ is the absorptivity of the left handed polarized light (as a function of the wavelength, λ), A_R is the absorptivity of the right handed polarized light (also as a function of the wavelength), $\varepsilon_L(\lambda)$ is the extinction coefficient of the molecule in response to the left-handed polarized light (a function of the wavelength, λ), $\varepsilon_R(\lambda)$ is the extinction coefficient of the molecule in response to the right-handed polarized light (a function of the wavelength (λ), b, is the pathlength, and c is the concentration of the sample. The ellipticity of a sample will vary according to the degree of absorption of left or right handed light, which is in turn defined by the extinction coefficient, which is in turn also defined by the wavelength. Thus CD spectra occur in locations of the electromagnetic spectrum wherever there also exists an absorption spectral band.

Some, but not all molecules exhibit optical activity. Most biological molecules, and DNA in particular, exhibit optical activity due either to the chiral carbons, such as the anomeric carbon of deoxyribose or to superasymetry about the structure, such as helicity of the DNA backbone. A circular dichroism signal is the result of many different factors contributing to the signal as whole. One such interaction or factor is the base stacking interactions. The orientation of the bases in space is such that the interaction of the pi-electrons is maximized, allowing the bases to stack on top of each other and interact with adjacent π-electron centers – the transitions from π to π* are closely linked and can interact to strongly absorb circularly polarized light. In addition, the deoxyribose of the DNA backbone is asymmetric and can give rise to an induced CD absorption between the sugar and the purine or pyrimidine base. The combination of the induced CD absorption, the strong interaction between adjacent bases, and the asymmetric distribution of the bases in space about the helical axis, gives rise to an intense and measurable circular dichroism signal upon interaction with left or right-handed polarized light.[35,120]

On a theoretical basis, CD spectra could be correlated to structural features or changes from quantum mechanical calculations alone. However, to date most attempts at structural prediction have proven to be insufficient for the determination of accurate structural models. Hence the role of CD and the interpretation of CD spectra has been reduced to the level of empiricism. Prior CD spectral information that has been correlated to a known structural feature can be correlated to new CD spectra based on the similarity between the reference (also called the "basis spectra") and observed CD spectra. While empirical spectra interpretation is a major limitation of CD spectroscopy,

much structural information can still be extracted from CD spectra. In fact CD spectroscopy is the only tool sensitive enough to detect subtle structural changes to DNA and proteins. Under defined conditions, the CD spectra can be used to follow transitions from one structural form of DNA to another without the existence of prior spectra. The structural form can be deduced from the solution conditions as a function of CD spectra. For example the percent duplex to single-strand DNA denaturation with respect to the solution temperature can be determined, in absence of a classic single-strand DNA CD spectra based solely on the deduction of extent of denaturation based on solution temperature. This is the basis of the CD-spectra as a function of temperature ("CD-Melts") and the CD spectra as a function of sodium concentration ("CD-Salts"). In the former, information about extent of denaturation with respect to temperature and the number of transition states was probed in absence of any defined "single-strand" CD spectra. The end-state at 90 ^{o}C was known from other techniques, and could be inferred on the basis of CD spectra alone. In the later examples, CD-salt, the information sought not just transition as function of salt, but lack of transition as function of salt. In this application, the nature of the DNA structure, or integrity of the DNA structure was investigated to determine what changes to the DNA structure occurred as sodium concentration increased, and at what concentration significant deviations from a classic B-form DNA this occurred.

Circular dichroism spectroscopy is sensitive, requires little sample and is nondestructive. This technique can be utilized to monitor structural transitions within defined limits. Since CD spectra are dependent on base interactions, any change to the nature of the

interaction, such as changes in the nucleotide orientation, changes to the twist or rise of the DNA molecule as well as the handedness of the DNA can be detected and followed by CD spectrophotometry. However, absolute structural interpretation of CD spectra is limited to either previously determined CD spectra, or structural features that have been correlated on the basis of other high-resolution techniques.

3.6.1.2 Sample Preparation and Procedures for CD Spectrophotometry

The DNA stock solution, in either water or low-salt PBS (5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH7.0), was diluted to a final strand concentration between 1 to 9 μM in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH7.0 from concentrated buffer stock solutions. The sodium chloride concentration was adjusted, after correction for the amount of sodium contributed by the buffering pair, by adding an appropriate aliquot of a 5M NaCl solution to each sample tube. In all cases the optical densities of the DNA solutions were between 0.75 and 1.25 absorbance units at 260 nm. The solution was heated to 90-100 °C for 20 minutes, then slowly cooled to 25 °C over 60 minutes.[121] The samples were equilibrated at 2-5 °C for 48 hours prior to measurement of the CD spectra.[122] For measurement of the CD spectra, the samples were scanned from 320 nm to 210 nm, with variable bandwidth, in 1 nm steps, at 25 °C, with a 3-second averaging time. Triplicate scans were averaged and then manually background corrected, adjusted for offset and smoothed with the DOS-based Aviv version 3.2 CD software. The corrected data were imported into Excel for conversion to molar ellipticity, and then plotted with SigmaPlot, version 5.0, Cricket Graph, or Prism version 2.01.

3.6.2 CD Melts

3.6.2.1 Essential Principles

The nature of the transition of the triplet-repeat DNA sequences as they denature from a hairpin to a single-stranded species was of interest not only for the purposes of defining the beginning and end states of the DNA structures, but for determining the number and types of transitions the DNA molecules undergo in accomplishing this transition. Because CD spectrophotometry is sensitive to both local and global changes in the base stacking interactions of the nucleotide bases, it was the method of choice to analyze the nature and number of states for the hairpin to single-strand transition of triplet repeat DNA sequences. To monitor these structural changes the CD spectra of the DNA samples were measured at increasing temperatures from 10 to 90 °C using a fine (1 nm) spectral step size to capture all spectral information.

3.6.2.2 Sample Preparation and Procedures for CD Melts

The DNA stock solution either in water or low-salt PBS (5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH7.0) was diluted to a final strand concentration between 1 to 9 μM in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH7.0, 100 mM NaCl using concentrated stock solutions. In all cases the optical densities of the DNA solutions were between 0.75 and 1.25 absorbance units at 260 nm. The solution was heated to 90-100 °C for 20 minutes, then slowly cooled to 25 °C over 60 minutes.[121] The samples were equilibrated at 2-5 °C for 48 hours prior to measurement of the CD spectra.[122] Prior to all CD measurement the samples were transferred from 2-5 °C storage and equilibrated at the start temperature in thermally equilibrated cuvettes for 15 – 20 minutes. The

cuvettes were stoppered with silicon-rubber stoppers to prevent evaporation and sample loss. The Aviv "T-write" macro generator was used to prepare instructions that computer-controlled all scan parameters including scan rate, ramp temperature, ramp step, equilibration time and slit bandwidth for all subsequent scans. The sample was scanned from 320 nm to 210 nm in 1nm steps with 1.5 nm bandwidth using single scans. The scan rate was 2 nm/min. The temperature step was 5 °C with an equilibration time of 10 minutes between subsequent scans. The start temperature was 10 °C and the end temperature was 90 °C. The thermal rate of change from one temperature to the next was 10 °C/min.

3.6.3 CD Spectra as a Function of [Sodium ion]

3.6.3.1 Rationale and Essential Principles

Among the many applications and strengths of CD spectrophotometry, this technique can be used to monitor any change the solution environment has on DNA structure, such as changes in the base-to-base spacing that is a reflection of changes in the phosphate backbone spacing. Changes in sodium concentration will alter the spacing of the phosphate backbone by shielding the repulsive negative charges of the phosphate groups. The higher the sodium concentration, the higher the counterion condensation, the more compact a structure the DNA can assume due to shielding of the phosphates along the DNA "backbone". The nature of the structures formed and the stability of those structures with respect to the salt environment was important not just for determination of the underlying structural reasons of the differential ion binding term, but also for the

kinetics studies as well where the sodium concentration increased from 10 mM sodium to 500 mM sodium. Thus these studies not only helped to define the structures but confirmed whether or not the beginning states of the kinetics samples were equivalent.

3.6.3.2 Sample Preparation and Procedures

The DNA stock solution, in either water or low-salt PBS (5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH7.0), was diluted to a final strand concentration from 1 to 9 μM in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH7.0 using concentrated stock solutions. The sodium chloride concentration was adjusted, after correction for the amount of sodium contributed by the buffer, by adding an appropriate aliquot of a 5M NaCl solution to each sample tube. In all cases the optical densities of the DNA solutions were between 0.75 and 1.25 absorbance units at 260 nm. The solution was heated to 90-100 °C for 20 minutes, then slowly cooled to 25 °C over 60 minutes. The samples were equilibrated at 2-5 °C for 48 hours prior to measurement of the CD spectra. All data were background corrected, offset and smoothed with the DOS-based Aviv version 3.2 CD software. The corrected data were imported into Excel for conversion to molar ellipticity, and then plotted with either SigmaPlot, version 5.0, Cricket Graph, or Prism version 2.01.

3.6.4 Native Gel Electrophoresis: Analysis of Structure

3.6.4.1 Essential Principles

Native Gel electrophoresis differs from denaturing gel electrophoresis (UREA-PAGE) with respect to the integrity of DNA secondary structure during separation. With native gel electrophoresis, the environmental solution conditions are predetermined to maintain

the secondary and other higher order structures of the DNA molecule. There are no denaturants and the samples are electrophoresed at low temperatures under mild solution conditions that can occasionally (but not always) mimic the physiological conditions of the cell. Because the secondary structure remains intact, the oligomers will migrate on the basis of both size and charge. This technique can therefore yield information with respect to the structures formed by DNA oligomers, duplexes, etc., specific to the environmental solution conditions selected. Consider a hypothetical sequence of DNA that can exist as two different secondary structures – hairpin and single stranded form. Though both forms have the same mass, the hairpin form has a smaller Stokes radius (smaller size) and less frictional resistance relative to the single-stranded form. Because the charge is equivalent for both species the hairpin form with the lower frictional resistance will migrate faster than the form with the larger frictional resistance (single strand). In this example the separation was caused by size, because both forms had similar net charges. However, alteration of the exposed charges can also change migration patterns. Consider for example, a DNA hairpin undergoes a hypothetical conformational change that made 50% of the phosphates inaccessible to the external solution. The electrophoretic pattern would be quite different – the two forms would nearly comigrate because the hairpin, while possessing a smaller size, has less charge and the single strand, while having more charge, had a larger size. The rate of migration in native gel electrophoresis is dependent upon the net charge of the DNA and the frictional resistance of the molecule in the support medium. Separation in native gel is not caused by size alone.

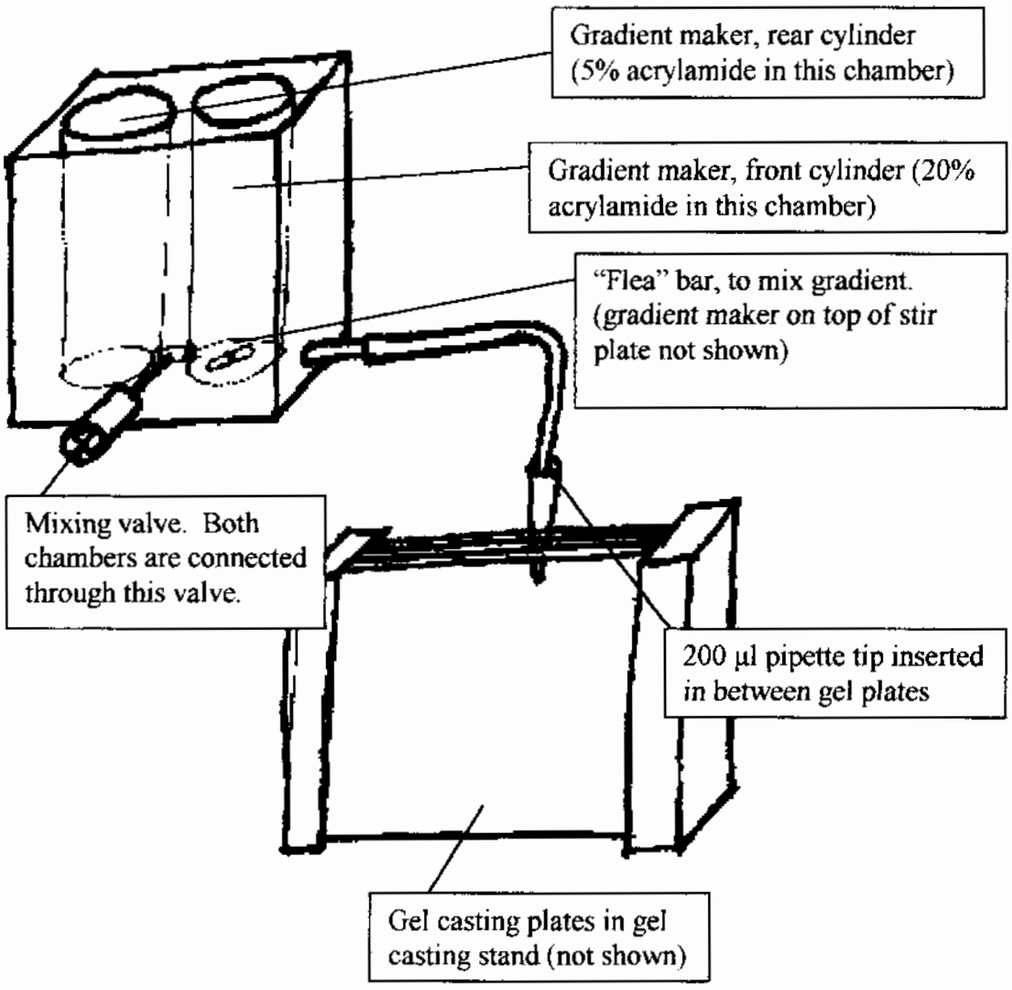
Native gel electrophoresis at a fixed concentration of acrylamide, while a powerful tool for the solution conformation of biologically relevant macromolecules, has some limitations. Because the gels are electrophoresed under low voltages, the bands are prone to diffusion and band broadening. In addition, structural variations within a single band (actually a small sub-population of molecules from the entire sample set) can further decrease resolution by providing a range of available states for a molecule to occupy. These different states will also have slightly different electrophoretic migration patterns producing, as the result, a band with diffuse edges. Because of this phenomena, the resolving power of fixed acrylamide native gels is somewhat limited. A special type of native gel – the polyacrylamide gradient or pore gradient gel avoids this problem, and is the only native gel technique suitable for size-analysis of biomolecules in their native state. In a pore-gradient gel, the polyacrylamide concentration increases from top to bottom, creating a gel where the pore size becomes progressively smaller from top to bottom.[123] This pore-gradient serves to focus the sample during separation, by restricting the migration of molecules to a specific size range within the gel. In practice, DNA (or proteins) will migrate through the gel until they reach a point where no further migration is possible due to restrictions of the pore size. Band focusing also serves to increase the sensitivity of this technique – samples that normally migrate as diffuse low intensity bands on a fixed percentage acrylamide gel can become focused high intensity bands with the pore-gradient gel technique.

3.6.4.2 Sample Preparation and Experimental Procedures

Samples for native gel electrophoresis were prepared from concentrated stock solutions to achieve a final DNA concentration between 100 – 200 μg/ml in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, 100 mM NaCl pH 7.0. The solutions were heated from 90 –100 °C for 10 minutes then slowly cooled to room temperature over 60 minutes and equilibrated at 4 °C for a minimum of 48 hours.[124] At the same time, 1 μl of molecular size markers (G/BRL 10 bp ladder) was loaded into the bottom of two 500 μl microcentrifuge tubes containing 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, 100 mM NaCl pH 7.0 and treated in a similar manner. The optimum range of acrylamide concentration for separation of oligomers from 12 to 75 nucleotides long was empirically found to be from 5 - 20% polyacrylamide. Approximately 25 ml of the 5% and 20% w/v acrylamide solutions were prepared from concentrated stock solutions in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, 100 mM NaCl pH 7.0 in 10% glycerol.[125] To 25 ml of acrylamide solution 75 μl of 10 % ammonium persulfate, and 30 μl of *N,N,N',N'* -tetrametheylethylenediamine (TEMED) was added and immediately mixed.[126] The 5% acrylamide gel solution was immediately poured into the rear chamber of the gradient mixer, and the 20% acrylamide solution was immediately poured into the front chamber of the gradient mixer (Figure 29). The acrylamide gradient was poured from 20% to 5% from the bottom of the gel casting stand to the top (Figure 29), then allowed to polymerize undisturbed with the sample comb in place. A 10 μl aliquot, representing approximately 1 - 2 μg of DNA was removed from the DNA solutions at 4 °C and equilibrated at 37 °C for a minimum of 30 minutes prior to further processing. To each sample solution, 10 μl of 50% glycerol, 0.1% bromophenol blue, 0.1% xylene cyanol, in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, 100 mM NaCl pH 7.0

Figure 29: The acrylamide gradient maker and gel casting stand.

The gradient-maker was a shop-made device consisting of two cylinders connected with a common valve to enable mixing, and a take-off valve to enable gel pouring. Several gradient pouring devices are available through commercial suppliers.



(SHU Native Gel Sample Buffer) thermally equilibrated at 25 °C was added to the DNA samples just prior to sample-loading of the electrophoresis gel. The samples were loaded onto the polyacrylamide gradient gel in a Hoeffer electrophoresis tank equilibrated at 37 °C and then electrophoresed at 75 V, constant voltage, 37 °C, with 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, 100 mM NaCl pH 7.0 electrophoresis buffer.[127] Upon completion, the gel was removed from the cassette, and immediately immersed in a polypropylene tray containing 100 ml of 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, 100 mM NaCl pH 7.0 with 10 µl of Syber-Green I staining dye.[128] The tray containing the gel and staining solution was covered with foil[129] then gently rotated at room temperature for 15 minutes.[130] The stained DNA bands were visualized with a UV-transilluminator and all densitometry an photography accomplished with an AlphaImager digital camera system. All gels were inspected for satisfactory resolution and detection as shown by the molecular weight markers. Gels producing unusual migration patterns ("smiles or frowns) faint bands, or poor resolution were repeated.[131]

3.6.5 [DNA] Dependence of T_m

3.6.5.1 Essential Principles

A unique feature of the Marky and Breslauer construct of the van't Hoff equation is the relationship between reaction molecularity, strand concentration and equilibrium constant for self-complementary sequences of DNA.[36] The molecularity of a transition can be determined with the following equation (Equation 1 , repeated below),

Equation 1: $$1/T_m = \{(n-1)R/\Delta H\}\ln C_t + [\Delta S-(n-1)R\ln 2+R\ln(n)]/\Delta H$$

where n is the molecularity of the association, C_t is the total strand concentration, R is the universal gas constant, and T_m is the transition melting temperature. A similar relationship was derived for non-self-complementary sequences of DNA, namely

Equation 23: $$1/Tm = [(n-1)R/\Delta H^{\circ}]\ lnC_t + [\Delta S^{\circ}-(n-1)Rln2n]/\Delta H^{\circ}$$

Where, again, n is the molecularity of the association, C_t is the total strand concentration, R is the universal gas constant, and T_m is the transition melting temperature. A plot of $1/T_m$ versus lnC_t yields a plot with slope of $-(n-1)R/\Delta H$ and intercept of essentially $\Delta S/\Delta H$ (note that the two equations differ from the intercept location due to the differences in the entropy terms). With these two equations, information with respect to the complementarity of a sequence of DNA, the transition enthalpy (for transitions with molecularity greater than 1) and the molecularity of transition is obtained. Unimolecular transitions – for example a change in structure from a self-complementary hairpin to a single-stranded species, will reduce the slopes of both equations to zero. Transitions with molecularities of two or more will produce lines with negative slopes. All other terms being equal, the greater the molecularity of the transitions, the more negative the slope. Therefore, analysis of a plot of $1/T_m$ versus lnCt can yield information with respect to the molecularity, and further clarify the type of structure formed by a segment of DNA. In addition, these plots can also serve as a useful measure of the extent of self-complementarity of a DNA sequence. Line fits with the self-complementary sequences DNA model will have a lower intercept than the same line fitted with a non-self-complementary model. Additional information about the thermodynamics of the transition can also be derived from this plot. In many instances, this type of data analysis is more accurate than the traditionally used curves fitting approach used for single data

point determinations of transition enthalpy. The slope of the line determines the transition enthalpy and is given by the first term of (Equation 23) whereas the intercept determines the transition entropy, and is given by the second term of these equations.

3.6.5.2 Sample Preparation and Procedures

Samples of DNA were prepared from concentrated stock solutions to achieve a starting DNA concentration of approximately 370 μg/ml in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, 100 mM NaCl, pH 7.0. The DNA samples were heated from 90 – 100 °C for 10 minutes, then cooled to room temperature over 30 minutes. The samples were degassed in a glass dessicator vessel (without dessicant) under vacuum for 15 minutes at 25 in Hg, then spun in a microcentrifuge at 14,000 rpm to pellet any particulates. The starting concentration of each oligomer solution was determined by the A260 spectral scan assay with an aliquot of the concentrated solution. The samples were transferred to marked cuvettes [132] and the cuvettes were tightly capped with 13D silicon rubber stoppers (Fisher Scientific).[133] The Cary model 100E Bio spectrophotometer was programmed with Cary Thermal software to heat the samples from 10 °C to 90°C at 0.5 °C/minute, then cool the samples from 90 °C to 10 °C both at 0.5 °C/minute with a 10-minute equilibration between ramps. The temperature "ramps" were repeated for a total of 8 –10 cycles while absorbance was monitored at the wavelength of maximum absorption difference for each strand (determined by the "UV-Thermal Melt Wavelength Optimization" protocol below). Most thermal melt scans were performed under continuous nitrogen purge of the sample chamber to prevent water condensation on the cuvette surfaces.[134] Thermal melts with significant hysteresis indicating nonequilibrium conditions or noisy melt profiles were repeated. Upon successful data collection of the

thermal melting profile at one concentration, the DNA solutions were recovered from the cuvette, and diluted with 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, 100 mM NaCl pH 7.0 to the next lowest DNA concentration in the measurement series. In all, the DNA samples were measured from 370 to 3.7 μg/ml in a melt-recover-dilute cyclical manner.

The melting points for all DNA sequences was determined with Varian Cary "thermal" software version 2.0 using a hyperchromicity model to fit the UV-melt data-set to a self-complementary model with n = 1. All melting temperatures (T_m), were entered onto an Excel for spreadsheet collation and calculation of the X-axis ($\ln C_t$) and Y-axis ($1/T_m$) data parameters. The data were then imported into Prism for graphical analysis. The slope of the data set was determined by least-squares linear regression analysis of the data set to within 10% error.

3.6.6 UV-Thermal Melts as a Function of [Salt]

3.6.6.1 Essential Principles – DNA Structure, T_m and the Differential Ion Binding Term.

The spacing or rise of the DNA helix is determined in part by the shielding of the repulsive negative charges of the DNA phosphate backbone. Counterion condensation between positively charged sodium ions in solution and the negatively charged phosphate moieties of the backbone can reduce this repulsive force, and allow the DNA helice to assume a more compact structure. The repulsive forces, however, contribute to DNA instability. Reduction of this force contributes to the overall stability of the DNA molecule and therefore hairpins containing mismatches facing the bulk solution, should

have less instability resulting from a phosphate repulsion (not to be confused with overall stability). Others factors, such as the stability of the base stacking interactions determine the overall stability. Salt effects can therefore be used as a secondary measure of the stability of a DNA structure. In addition salt effects can indirectly measure the water activity, and the relative phosphate-spacing between different structural forms through determination of the differential ion binding term. The differential ion binding term, n, gives information about the number of ions, in this case sodium ions, released upon denaturation. Because the sodium ion concentration determines in part the spacing of the phosphate groups, the number of sodium ions released gives an approximate measure of the spacing of the phosphates along the DNA backbone. Double stranded DNA should have more compact structures, in terms of the helical rise, than triplet repeat hairpin structures that contain a mismatch every third nucleotide.

The theoretical basis for this measurement was described originally by Manning[135] and Record[136] who developed the polyelectrolyte theories, and derived an equation describing the relationship between [salt] and T_m (Equation 24),

Equation 24: $$dT_m/d\log[M] = \{2.303RT_m^2/\Delta H_m\}\Delta n$$

where T_m is the midpoint of the melting transition, [M] is the molar salt concentration, R is the universal gas constant, ΔH is the transition enthalpy (assumed to be independent of the Na^+ concentration), and n is the differential ion binding term. The differential ion binding term allows calculation of the number of sodium ions released upon strand denaturation. This term will serve to indicate whether longer repeats release more cation per nucleotide than shorter repeats.

3.6.6.2 Sample Preparation and Procedures for UV-Thermal Melts as a Function of [Sodium Ion]

Samples of DNA were prepared from concentrated stock solutions to achieve a starting DNA concentration between 27 - 37 μg/ml in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH 7.0 with sodium concentrations spanning the range of 15 mM to 1000 mM NaCl. The DNA samples were heated from 90 – 100 °C for 10 minutes, cooled to room temperature over 30 minutes, then degassed with a Savant Speed Vac under vacuum for 10 minutes.[137] The samples were spun in a microcentrifuge at 14,000 rpm (to pellet any particulates), transferred to marked cuvettes, [132] and the cuvettes were tightly capped with silicon rubber stoppers (Fisher Scientific).[133] A Cary model 100E Bio spectrophotometer with Cary Thermal software was programmed to heat the samples from 10 °C to 90 °C at 0.5 °C/minute, then cool the samples from 90 °C to 10 °C both at 0.5 °C/minute with a 10-minute equilibration between ramps. The temperature "ramps" were repeated for a total of 6 - 8 cycles while absorbance was monitored at the wavelength of maximum absorption difference for each strand (determined by the "UV-Thermal Melt Wavelength Optimization" protocol below). Most thermal melt scans were performed under continuous nitrogen purge of the sample chamber to prevent water condensation on the cuvette surfaces.[134] Thermal melts with significant hysteresis indicating nonequilibrium conditions or noisy melt profiles were repeated.

Acceptable data was processed with the Varian "Thermal" software version 2.0 using the total strand concentration, the molecularity (n = 1 for HP to SS transitions; n = 2 for DUP to SS transitions) and a self-complementary model to fit the data for T_m, enthalpy

and the equilibrium constant. The free energy and entropy were automatically derived using the same software from the equilibrium constant and the Gibbs equation. All data were entered onto an Excel spreadsheet for collation and further data processing. The data were imported into Prism graphical analysis software, and a plot of T_m versus log[sodium] was constructed. The slope of the line was fit to a least squares linear regression model. The differential ion binding term, was calculated from the slopes of these lines with (Equation 25):

Equation 25: $T_m = [\{2.303RT_m^2/\Delta H_m\}\Delta n]\ [\log [M]]$

where the X-axis is log[sodium], Y is the melting temperature T_m, R is the universal gas constant, and n is the differential ion binding term.

3.7 Thermodynamic Characterization

3.7.1 UV-Thermal Melt Wavelength Optimization

3.7.1.1 Essential Principles of Wavelength Optimization

The wavelength used to monitor DNA transitions is a critical element in the acquisition of good DNA melting curves. Monitoring of a DNA transition at the incorrect wavelength can result in little or no apparent melting profile when in fact denaturation is occurring with the DNA sample.[138] The extinction coefficient of DNA is dependent upon the nature of DNA sequence, and the extent of denaturation of the DNA molecule. "Unexposed" nucleotides absorb less light (lower extinction coefficient) than unexposed nucleotides due to base stacking interactions between the bases. It is therefore prudent, before undertaking any DNA melting experiment, to measure the absorbance spectrum of

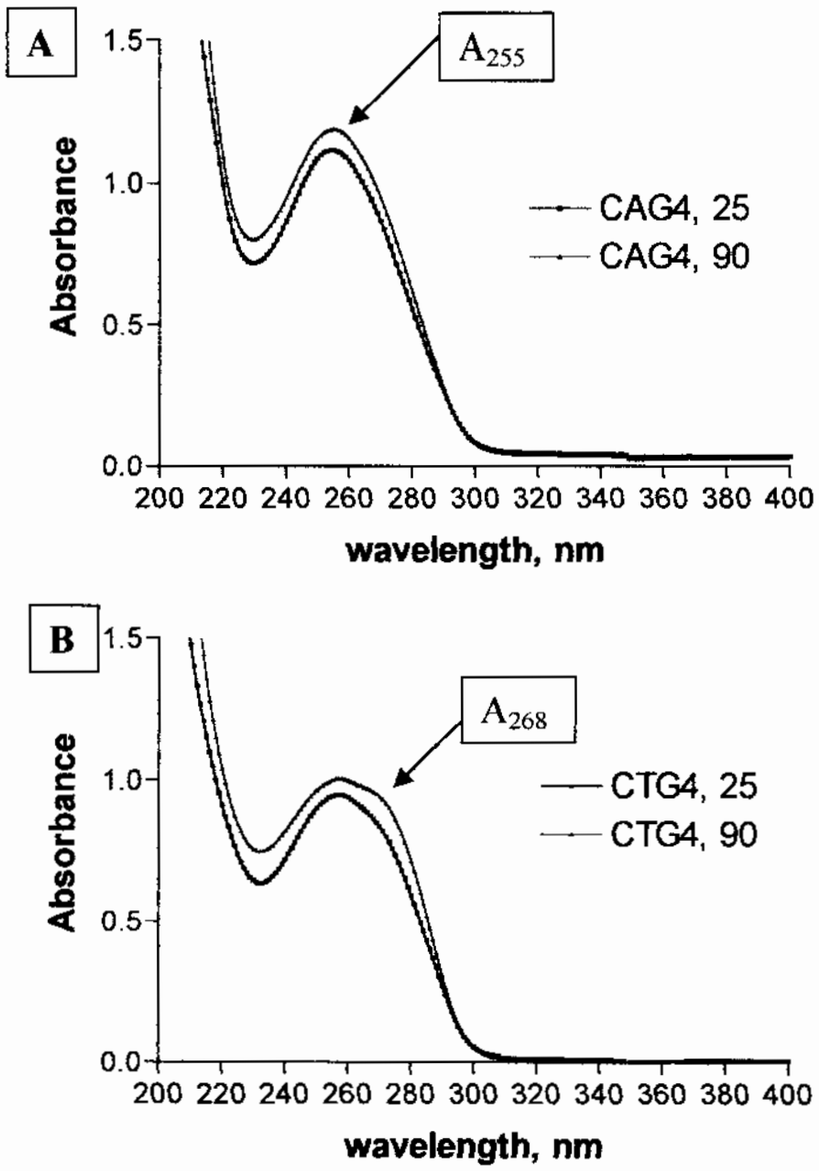
a DNA sample under fully denatured and fully folded conditions. The proper wavelength to monitor the DNA melting experiment is the wavelength with both the maximum signal and the maximum difference between the folded and denatured form of the DNA molecule.

3.7.1.2 Sample Preparation and Procedures for Wavelength Optimization

Samples of DNA were prepared from concentrated stock solutions to achieve solution conditions relevant to the experimental parameter under observation. The DNA samples were heated from 90 – 100 °C for 10 minutes, cooled to room temperature over 30 minutes, then degassed with either a Savant Speed Vac for 10 minutes[139] or a glass dessicator (without dessicant) under vacuum at 25 in Hg vacuum. The samples were spun in a microcentrifuge at 14,000 rpm (to pellet any particulates), transferred to marked cuvettes, [132] and the cuvettes tightly capped with silicon rubber stoppers (Fisher Scientific).[133] A Cary model 100E Bio spectrophotometer with Cary "Scan" software was programmed to equilibrate the DNA solutions at 10 °C (or other start temperature) for 15 minutes then scan the sample from 400 nm to 200 nm with 1 nm bandpass at 200 nm/minute. Upon successful data collection the spectrophotometer was programmed to equilibrate the DNA solutions at 90 °C (or other final temperature) for 15 minutes then scan the sample from 400 nm to 200 nm with 1 nm bandpass at 200 nm/minute. The spectra from the two temperature ranges were overlaid and the wavelength displaying both the largest signal and the largest difference between the two temperatures was noted (Figure 30). The wavelength noted was used to monitor the DNA thermal denaturation experiment.

Figure 30: A UV-difference spectra.

A. Sample spectra of a CAG4 oligomer scanned at 25 °C (black, lower scan) and 90 °C (red, upper scan). The optimum wavelength for monitoring the DNA hyperchromicity transition, indicated by arrow below is 255 nm. **B.** Sample spectra of a CTG4 oligomer scanned at 25 °C (black, lower scan) and 90 °C (red, upper scan). The optimum wavelength for monitoring the DNA hyperchromicity transition, indicated by arrow below is 268 nm. Note that the optimum wavelength for UV-melt analysis is not necessarily the wavelength of maximum absorbance.



3.7.2 UV-Thermal Melts

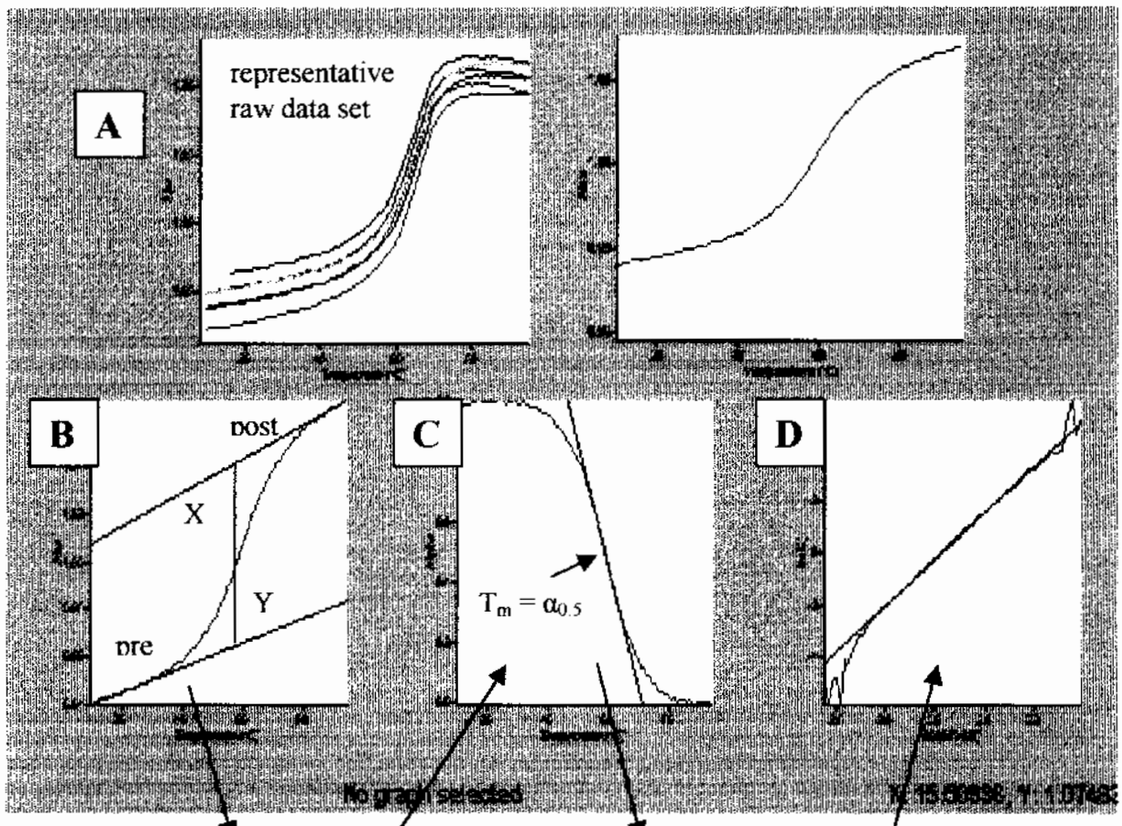
3.7.2.1 Essential Principles, UV-Thermal Melt

The extinction coefficient of DNA will vary according to both the sequence content and the solution environment of the DNA strand. Changes in temperature will disrupt the base stacking interactions that stabilize the DNA helices and render the bases more accessible to light energy. This results in an increase in the extinction as DNA is melted from a stably base-stacked structure. The increase in absorbance as the DNA progresses from some structured form to a single-stranded form is called hyperchromism, in reference to the fact that absorbance increases on melting. Hyperchromism can be utilized to monitor the melting transition of DNA from a native form to a single-single stranded species, or, in the reverse direction, the association of two single strands into a stable duplex.

The characteristic s-shaped melting curve (Figure 31) for DNA melting can be utilized to determine the melting temperature, the composition of the equilibrium populations at any point of the transition, the enthalpy of dissociation, the equilibrium constant and by secondary calculations, the free energy and entropy terms for the transition (Figure 31). Melting of DNA can be thought of as a thermodynamically-driven process with two end-states consisting of either the fully folded native form or the single-stranded denatured form. With this model, the melting of DNA involves various equilibrium distributions of these two species that are a function of the denaturation conditions. If the denaturant is

Figure 31: Data processing of a UV-melt experiment.

The principles of this operation are automatically performed with the Varian "Thermal" software, and were first described by Marky and Breslauer. **A.** A single scan is selected from the raw data set for analysis (plot is absorbance vs temperature). **B.** The raw data (red) is characterized by defining the pre-transition and post-transition baselines (black) marking the native and single-stranded states of the DNA respectively (absorbance vs. temperature). **C.** A mathematical transform is used to convert the raw data (red) into an α-plot, that describes the extent of reaction between the two end-states (alpha vs. temperature). A fit (black) is drawn to the linear region of the transition for determination of the equilibrium constant as a function of temperature. **D.** The alpha plot is used to construct a lnK vs. 1/T plot, that defines the van't Hoff enthalpy (lnK vs. (1000)(1/T)).



The absorbance vs. T data is converted into the extent of transition, α,

$$\alpha = X/(X+Y)$$

and is used to construct an α plot (α vs T). The melting temperature, T_m, is defined as α = 0.5.

The α-term is used to derive the equilibrium constant, K, at each temperature using:

$$K = \alpha / [(C_t/n)^{n-1} (1- \alpha)^n]$$

A van't Hoff plot (ln K vs. 1/T) is constructed to determine the transition enthalpy, and entropy from :

$$\ln K = -(\Delta H/RT) + (\Delta S/R)$$

temperature then the distribution as reflected in the equilibrium constant, K, of each species is described by the van't Hoff relationship[36]

Equation 26: $\mathbf{d(\ln K)/dT = \Delta H/RT^2}$

From this relationship, Marky and Breslauer redefined the equation in experimentally measurable variables, such as the extent of transition, α (derived from absorbance data as shown in Figure 31), the molecularity of the association, n, and the strand concentration, C_t.[36]

Equation 27: $\mathbf{K = \alpha / [(C_t/n)^{n-1} (1-\alpha)^n]}$ **(nonself-complementary sequences)**

Equation 28: $\mathbf{K = \alpha / [n(C_t)^{n-1} (1-\alpha)^n]}$ **(self-complementary sequences)**

From this equation the van't Hoff enthalpy can be calculated by substituting for K into the van't Hoff equation, differentiating with respect to T then solving for enthalpy to yield:

Equation 29: $\mathbf{H_{VH} = (2 + 2n)R(T_m)^2(\delta\alpha/\delta T)}$ **(valid for both types of sequences)**

Another approach to data analysis called "curve analysis" determines the transition enthalpy from the shape of the alpha plot. In this type of analysis, the equilibrium constant, K, is calculated from α at each temperature (Equation 27 or Equation 28). The enthalpy is derived from a plot of lnK vs 1/T, with the following equation (Equation 30),

Equation 30: $\mathbf{\ln K = -(\Delta H/RT) + (\Delta S/R)}$

where K is the α-derived equilibrium constant, R is the universal gas constant, ΔH is the van't Hoff transition enthalpy, and ΔS is the transition entropy. The slope this plot gives ΔH/R and the intercept gives, ΔS/R. An alternative method of determining the transition enthalpy measures the concentration dependence of the melting transition (discussed in

section titled [DNA] versus T_m). Both approaches to obtaining thermodynamic information from denaturation experiments assume, as mentioned previously, that the DNA transition occurs in a two-state manner. At any point along the characteristic s-shaped curve, the population of DNA is divided, according to the α-term, into two distinct populations – native or denatured. Any deviation from this assumption, such as the formation of an intermediate structure, invalidates the use of this model because the composition of the population cannot be defined. If the population cannot be defined, then the equilibrium constant, or extent of reaction as a function of temperature becomes meaningless. The second assumption made by this model is that DNA denaturation is not accompanied by a change in heat capacity between the native and denatured forms (Equation 31).

Equation 31: $\Delta C_p = 0$

With a DSC thermogram, any changes in heat capacity are readily visible as relative changes in the height of the pre-transition and post-transition baselines. With UV-melt techniques, there is no equivalent measure of residual heat capacity that indicates such a change (a serious drawback of UV melting techniques). Because the enthalpy is dependent upon the heat capacity of the sample, and the heat capacity is determined over a range of temperatures, a change in heat capacity means that the enthalpy is not independent of temperature.

Through the use of UV-melting techniques, model-dependent thermodynamic information can be derived. This information can be subsequently compared to the calorimetrically derived thermodynamic data to determine if the transitions occur in a

two-state manner. In addition, the UV-melting techniques provide a secondary means to establish relative stability ranking for the sequences under consideration, is more sensitive than calorimetry and requires less sample than calorimetry. In addition this technique, like calorimetry, is non destructive. Drawbacks of this technique, include the fact that only DNA or DNA-related transitions can be monitored.

3.7.2.2 Sample Preparation and Procedures

Samples of DNA were prepared from purified, characterized and highly concentrated stock solutions to achieve conditions in solution that were relevant to the experimental parameter under observation. The DNA samples were heated from 90 – 100 °C for 10 minutes, cooled to room temperature over 30 minutes, then degassed for 10 minutes with either a Savant Speed Vac[139] or glass dessicator (without dessicant) under vacuum at 25 in Hg vacuum. The samples were spun in a microcentrifuge at 14,000 rpm (to pellet any particulates), transferred to marked cuvettes, [132] and capped tightly with silicon rubber stoppers (Fisher Scientific).[133] A Cary model 100E Bio spectrophotometer with Cary, "Thermal" software was programmed to heat the samples from 10 °C to 90°C at 0.5 °C/minute, then cool the samples from 90 °C to 10 °C at 0.5 °C/minute with a 10-minute equilibration between ramps. The temperature "ramps" were repeated for a total of 6-10 cycles depending upon the required accuracy of the measurement.[140] The absorbance at the wavelength of maximum absorption difference for each strand (see "UV-Thermal Melt Wavelength Optimization" protocol) was collected as a function of temperature. Thermal monitoring of the sample temperature was accomplished with either the Peltier block probe-set[141] or an external temperature probe inserted directly inside of the cuvette.[142] Most thermal melt scans were performed under continuous nitrogen purge of

the sample chamber to prevent water condensation on the cuvette surfaces.[134] Thermal melts with significant hysteresis (nonreproducible melting profiles in the same direction or in either direction, shifts in T_m etc) indicating nonequilibrium melting conditions, or thermal melts with noisy melt profiles were repeated with appropriate corrections to the experimental protocol.

Acceptable data, defined as spectra with low noise, good scan to scan reproducibility, and single transitions (Figure 32), were processed with the Varian "Thermal" software version 2.0 using the total strand concentration, the molecularity (n = 1 for HP to SS transitions; n = 2 for DUP to SS transitions) and a self-complementary model to fit the data for T_m, enthalpy and the equilibrium constant. All thermodynamic parameters were derived from an alpha-curve plot generated within the Varian "Thermal" software and were subject to the accuracy of the manually determined pre-transition and post-transition transition baselines. The free energy and entropy were automatically derived using the same software from the equilibrium constant and the Gibbs equation. All data were entered onto an Excel spreadsheet for collation and further data processing.

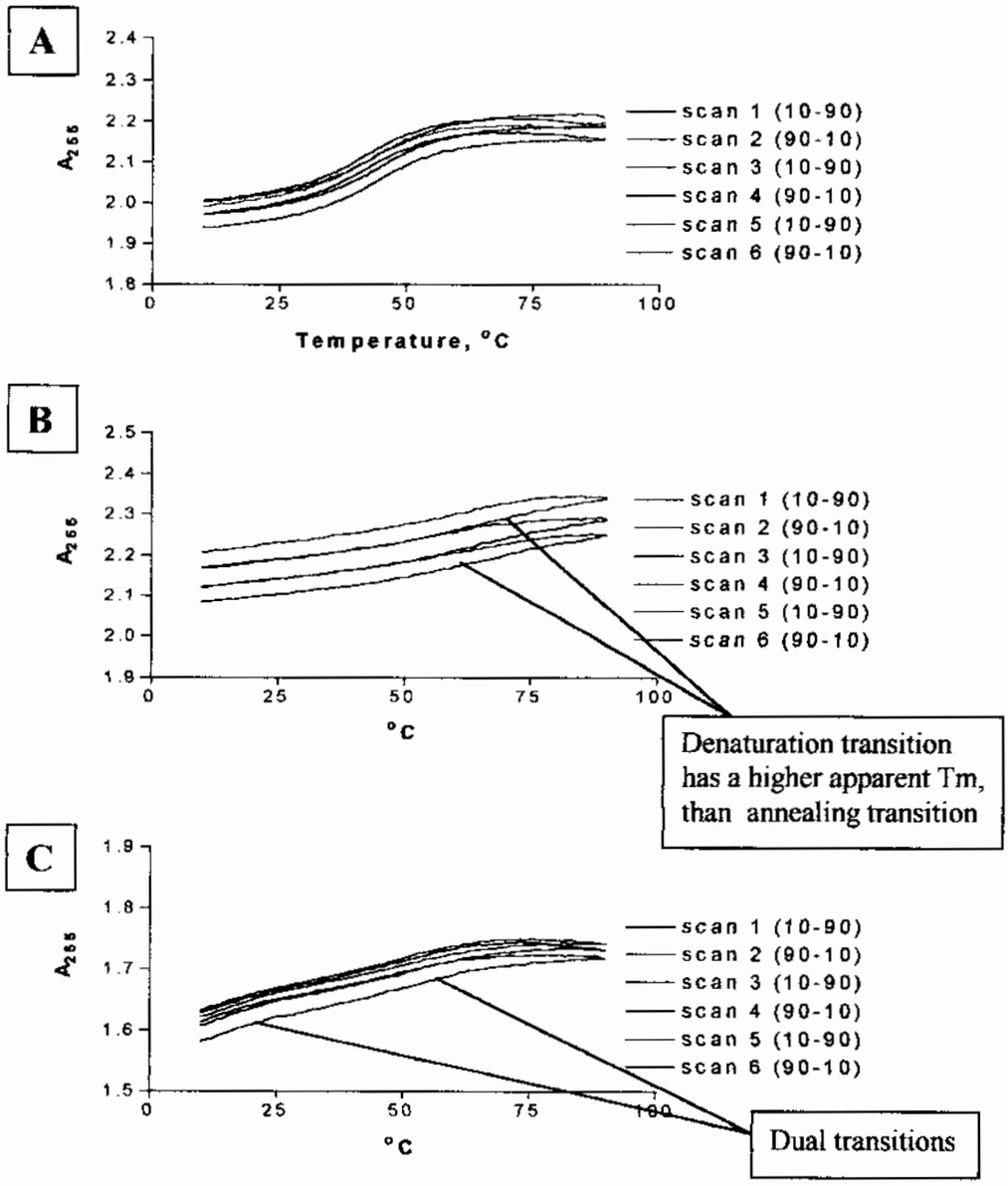
3.7.3 Differential Scanning Calorimetry

3.7.3.1 Essential Principles, and Challenges for DNA DSC.

The instrumentation for calorimetry essentially is one of three different classes – the batch calorimeter, the flow calorimeter and the titration calorimeter.[143] Flow calorimeters measure heat as a function of reaction time, and titration calorimeters measure heat as function of added reactant at some concentration. Batch calorimeters, which include the

Figure 32: A sample UV-melt experiment.

For each data set below there are six thermal scans – three in the "forward" direction from 10 °C to 95 °C and three in the "reverse" direction from 95 °C to 25 °C. **A**. A sample UV-melt measurement displaying good reproducibility with low noise. **B**. A UV-melt experiment showing hysteresis – note the lack of reproducibility from scan to scan. **C**. A sample UV-melt showing poor or secondary transitions.



differential scanning calorimeter, measure heat as a function of a thermodynamically relevant variable according to the instrument design. For example a pressure-peturbation calorimeter measures the heat releases as a function of pressure perturbations, which is in turn related to the volume change on transition. A differential scanning calorimeter measures the heat of transition, as function of temperature, as molecules undergo transitions from one structural form to another. Differential scanning calorimeters are of two basic designs – the heat flux or the power compensation calorimeter. The Nano-II DSC is a power compensation calorimeter. Heat flux calorimeters raise the temperature of the reaction cell and measure the heat transferred to the surroundings, or transferred from the surroundings. These types of calorimeters require sophisticated detection circuitry, as well as excellent heat isolation to avoid thermal transfer from the external surroundings to the measurement cell. Power compensation calorimeters measure the heat of transition as a function of temperature by measuring the amount of power required to maintain a constant temperature between a sample cell containing the molecule of interest plus buffer and a blank cell containing the buffer solution alone. As the temperature of the cell-block is slowly raised, according the user-specified program the macromolecules in the sample cell will undergo transitions with an associated release or uptake of heat. Such transitions can include refolding, changes in sugar orientation, changes in rise, changes in hydration, changes in structural form – all will have an associated heat event. Endothermic transitions will cause the temperature of the sample cell to lag behind that of the reference cell, because some of the energy in heating is used to effect the transition. Exothermic transitions will cause the sample cell temperature to run hotter than the reference cell. The amount of heat, in the form of electrical power

supplied from or subtracted by the external Peltier device to maintain a constant temperature between the sample cell and the blank cell is measured and reported as the μW of power required as function of temperature – the raw excess heat capacity data. This data plot, called the excess heat capacity plot, represents the amount of heat supplied to the sample cell that is above the heat required to raise the temperature of the cell-block. The raw heat capacity data is converted to the molar heat capacity using the molecular weight, cell volume and concentration of sample within the sample cell.

The theoretical framework that allows extraction of thermodynamic data from calorimetric measurements begins with the fundamental enthalpy relationship, where the enthalpy, H, is related to the energy of the system, E, and the work done by the system, PV (Equation 32).[144]

Equation 32: $$H = E + PV$$

The enthalpy is not know in absolute terms, only the changes in enthalpy are measured, reducing this equation to

Equation 33: $$\Delta H = \Delta E + \Delta(PV)$$

If this equation (Equation 33) is differentiated, we obtain the following relationship defined in terms of energy (E), pressure (P), and volume (V) (Equation 34).

Equation 34: $$dH = dE + PdV + VdP$$

However, the energy term, "dE" is related to the heat and work with the following relationship (Equation 35).

Equation 35: $dE = dQ + dW$

and substituting this equation (Equation 35) into the differentiated definition for enthalpy (Equation 34) produces the following relationship between heat (Q), work (W), pressure (P) and volume (V).

Equation 36: $dH = dQ + dW + PdV + VdP$

This equation (Equation 36) can be simplified with some restrictions or assumptions. First, the process is carried out at constant pressure (Equation 37)

Equation 37: $dP = 0$

This assumption accurately reflects the case in a calorimeter cell, where the applied pressure is held constant at 3 atmospheres. Second the only work allowed is mechanical work against the ambient pressure. Using the convention that work from the system to the surroundings is negative will produce the following relationship (Equation 38).

Equation 38: $dW = -PdV$

Substituting both the constant pressure (Equation 37) and work (Equation 38) relationships into the enthalpy definition produces the following relationship with respect to enthalpy and heat (Equation 39)

Equation 39:
$$dH_p = dQ_p - PdV + PdV + VdP$$
$$dH_p = dQ_p + 0 + 0$$
$$dH_p = dQ_p$$

To convert the heat term, Q, to experimentally and calorimetrically relevant parameters, requires another set of manipulations. Using the relationship between the heat (Q), temperature and the heat capacity (C_p) at constant pressure (Equation 40) ,

Equation 40: $C_p = dQ/dT$ **(definition)**

and solving this equation for dQ produces a relationship that can be substituted into the enthalpy/heat relationship (Equation 39) to produce the following equation (Equation 41).

Equation 41: $dH = C_p dT$

Integration of this equation gives (Equation 42), which is the equation that makes calorimetry possible.

Equation 42: $\Delta H = \int C_p dT$

Power data from a typical DSC experiment is converted to the molar excess heat capacity using the following relationship (Equation 43) where Q_p is the excess heat at constant pressure, T is the temperature, and M is the number of moles of sample.[145]

Equation 43: $(dQ_p/dT)(1/M) = C_p^m$

The applied excess power to the sample cell is converted to heat using the following relationship (Equation 44) where the heat (W) in μJoules, is related to the current (i) in μamperes, the resistance (R), and the time (t).[146]

Equation 44: $W = i^2Rt$

This equation (Equation 44) is a variation of Joule's Law (Equation 45),

Equation 45: $P = i^2R$

where a time element has been inserted to convert the power (P) in watts or Joules/sec into Joules.

With respect to UV-melting techniques the measurement of thermograms for DSC experiments requires a large amount (μmole) of highly concentrated sample. The typical range of concentrations required for a 30-mer oligonucleotide is, at a minimum, between 40 – 60 μM DNA (strand concentration) in 1.0 ml or about 400 - 555 μg/ml of DNA. Longer DNA sequences will require less sample (in terms of strand concentration) due not only to the molar increase in nucleotides, but to the formation of complex structures that contribute to the heat of transition. To place this quantity into perspective, a typical 1 μ-mole DNA synthesis yields approximately 1000 to 1500 μg of crude DNA. After purification, the overall yield will decrease, and depending upon the method of purification will produce approximately 800 to 1000 μg of DNA if purified by PolyPak II, and 500 to 750 μg of DNA if purified by RPHPLC. The total amount of full-length DNA derived from a high-yield 1-μmole DNA synthesis, that has been purified to high purity (>98%), in high yield (or else no signal) is required for one replicate of a DSC experiment. In spite of this limitation, DSC does offer significant advantages over other techniques, such as UV-melt. First, calorimeters are not dependent upon experimentally derived variables, such as an absorbance difference, that may or may not change as a function of denaturation. [147] DSC measures the heat output or intake directly. Any event that causes a heat signal, such as ligand release, structural transformations, or aggregation, will be detected by DSC. Because of the heat is measured directly, and the

thermodynamic parameters are determined from these measurements, calorimetry offers model independent means of assessing thermodynamic information. In addition, DSC offers the ability to accurately determine if the enthalpy is independent of temperature by examination of the pre-transition and post-transition baselines. Such analysis, and the accuracy of the enthalpy derived, is not possible with other techniques, such as UV-melt.

3.7.3.2 Sample Preparation and Procedures

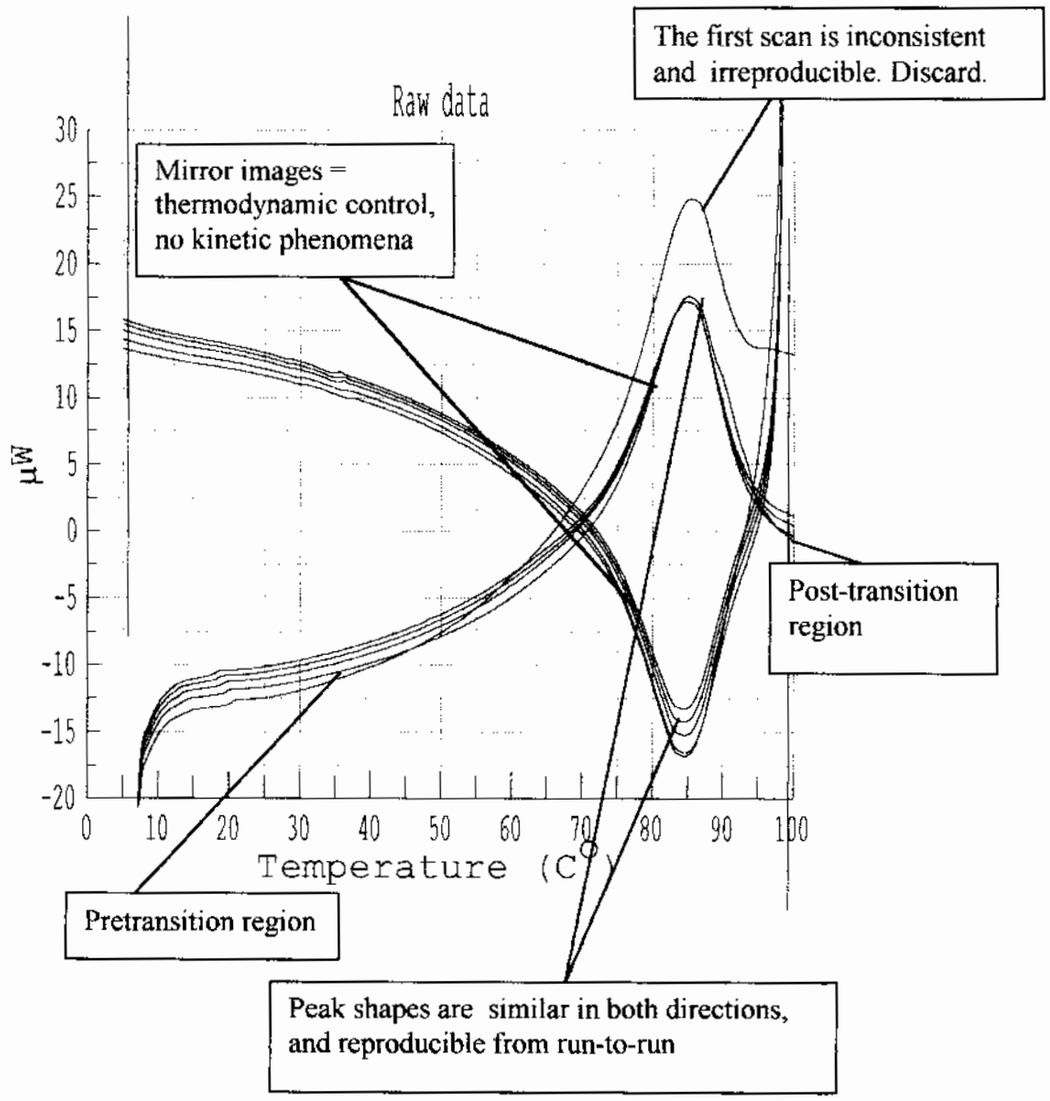
The buffer solutions used for all DSC-related procedures, including dialysis, PolyPak II lyophilate reconstitution, or cell rinsing, were characterized by DSC for residual contaminants that could affect sample measurement. This measurement not only qualified the buffer for use with DSC samples, but also served as the "blank" scan for subtraction of the residual heat capacity of the buffer. Buffer solutions with any peaks or baseline drifts larger than 5 μW were discarded. Because the nature of the buffer solution could change upon storage (bacterial growth, shifts in pH etc.) buffer versus buffer measurements were conducted, at a minimum, every fifth sample on the DSC. Samples for DSC were either exhaustively dialyzed into the experimentally relevant buffer solution or were derived from a PolyPak II purification process where the DNA was desalted and exhaustively washed on column, eluted and reconstituted in the experimentally relevant buffer. For most DNA samples in this project, the sample buffer was 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, 100 mM NaCl pH 7.0. Samples for DSC had to be highly concentrated because of the relatively low heat capacity of DNA oligomers (relative to proteins). The DNA samples for DSC were between 400 and 555 μg/ml (minimum concentration) in a volume from 750 to 1000 μl.[148] The beginning DNA concentration of each oligomer was measured with the A260 scan assay and any

solution with a concentration below 400 µg/ml was not used for DSC. The DNA sample solution was heated from 90 °C to 100 °C for 10 –15 minutes, mixed with a vortex mixer, then rapidly cooled to room temperature by centrifugation in a microcentrifuge for 5 minutes at 14,000 rpm. The DNA solution was filtered through a 0.45 µm nylon syringe filter then degassed in a glass dessicator (without dessicant) for 10 – 15 minutes at room temperature under a vacuum at 25 in Hg. The DSC capillary cells were thoroughly rinsed with duplicate 1 liter washes of deionized distilled water between samples, and then given a final rinse with triplicate 1 ml washes of buffer solution in each cell prior to sample loading.[149] The capillary cell of the reference side was loaded with 0.45-µm filtered and degassed buffer and the sample capillary cell was loaded with the filtered degassed sample. Bubbles within the cells were dislodged by slowly "pumping" the solution within the cell several times.[150] Because 750 µl of solution was required to load the cells, and the cell volume was 299 µl, there was a considerable amount of solution remaining after sample loading. The remaining sample was retained for subsequent DNA concentration determinations ("pre-DSC" DNA concentration sample). The solutions were allowed to equilibrate inside the DSC until the applied power between the cells reached a constant value over 15 minutes.[151] The cells were pressurized to 3 atmospheres of pressure to prevent evaporation and boiling of the sample. The Nano-II DSC was programmed to heat the samples ("ramp up") from 5 °C to 90°C at 1.0 °C/minute, then cool the samples ("ramp down") from 90 °C to 5 °C at 1.0 °C/minute, for 8 to 10 ramps with a 10-minute equilibration between ramps. All data were collected with "RunDSC" software and initially processed with "CpCalc" software from CSC Inc.

Post-measurement all thermographic raw data were initially inspected for completeness of the transition, overall shape of the transition, noise level, the number of peaks, the nature of the peaks, the amount of heat required for the transition (peak height), scan-to-scan reproducibility, heating/cooling symmetry, and vertical position of the DSC trace (Figure 33). Those data sets with incomplete transitions, asymmetric transitions, nonsymmetric heating/cooling profiles, and profiles with poor reproducibility (irreproducible shape or vertical position) were rejected and, if necessary, the thermal transition of the sample was remeasured with appropriate corrections. Samples that presented problematic thermographic profiles on reanalysis were discarded unless the duplicate sample reproduced the prior result. Acceptable thermographic data, defined as data sets or subsets with symmetric, reproducible peaks, with full pre-transition and post-transition baselines, and reproducible heating/cooling symmetry were imported into "CpCalc" software for blank subtraction and calculation of the molar heat capacities. Upon successful data collection, the sample was removed from the DSC and the concentration of both the pre-DSC and post-DSC samples were determined with the A260 scan assay and the summed extinction coefficient (Figure 26). Occasionally the DNA concentration was observed to increase as the result of evaporation or decrease as the result of buffer droplets within the sample cell. Data for those samples with concentration changes greater than 10% were rejected and the DSC experiment restarted. The DNA concentration used for all calculations was the post-DSC DNA concentration. The contribution of the buffer heat capacity to the DNA heat capacity measurement was subtracted in one of two ways, depending upon the experimental result. If all thermograms of a buffer blank were super-imposable then a representative blank scan

Figure 33: A sample DSC thermogram.

The melting transition for (CGG)15 in 10 mM sodium phosphate, 0.1 mM EDTA, 100 mM NaCl, pH7.0. The ramp rate was 1.0 °C/min, 8 ramps, starting in the heating direction, with a 10-minute equilibration between ramps. All thermograms are initially inspected for reproducibility, extent of transition, the nature of the pre-transition and post-transition baselines, and the overall response in terms of heat capacity.



was selected and subtracted from the heat capacity measurements of all scans of a particular sample. If the thermograms of the blank were not super-imposable across all scans then the individual thermograms were subtracted in a one-to-one correspondence with the sample scans - buffer blank 1 scan was subtracted from sample scan 1 etc. Thermodynamic information was calculated from the integrated area of the transition curve. The baselines of these transitions were drawn to match, as closely as possible, the slope and/or curvature of the pre/post transition baselines (Figure 34). Selection of the start and stop points for baseline fitting was consistently applied across all scans of a sample, and as much as possible, across all scans of a triplet repeat series. The molar heat capacity, the enthalpy, and the entropy were calculated from the integrated area under the heat capacity curve, the molecular weight in kD, the cell volume and the DNA concentration in mg/ml. The calorimetrically-derived van'tHoff modeling of the data was achieved with either Origin version 6.0 software or CpCalc software.

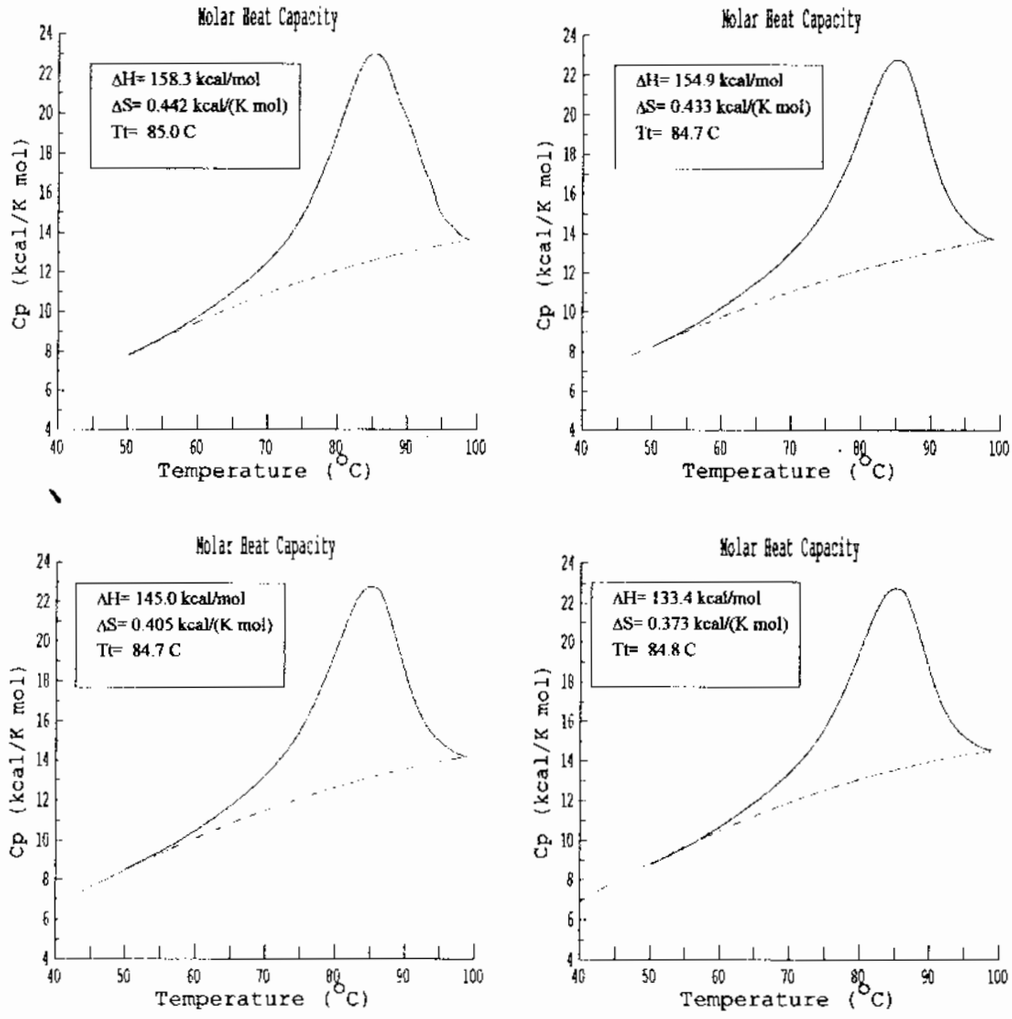
3.7.4 Calorimetric Data Modelling

3.7.4.1 Essential Principles and Relevance to UV-thermal Melts

Differential scanning calorimetry was used to detect structure to single-strand transitions for all oligonucleotides used throughout this project. With calorimetrically derived data, the area under the excess heat capacity curve, or Cp is used to determine the transition enthalpy as previously described. Integration of calorimetric data is not influenced by the shape of the curve. However, calorimetric enthalpy and entropy is not the only information that can be obtained from a thermogram. Like the UV-melting curves, the shape of the excess heat capacity plot enables calculation of the theoretical, or van't Hoff

Figure 34: An integrated DSC thermogram.

The pre-transition and post-transition baseline (black) is drawn to match, as closely as possible, the slope and curvature of the raw data (blue). The start and stop points are consistently applied to samples within a dataset. A polynomial fit was applied from 50 – 90 °C to four thermograms in the endothermic (heating) direction from the CGG15 data set (above).



transition enthalpy or enthalpies for the process under consideration. As with modeling of UV-melt absorbance versus temperature experiments the shape of the modeled curve determines, in large part, the derived van't Hoff enthalpy. Because of this fact, due diligence must be exercised in the drawing of the model data curves. Like UV-melts, modeling of a calorimetric thermogram is subject to many of the same constraints and problems associated with UV-melts. Such problems, in general, are related to the accuracy the modeled curve describes the calorimetric or "real" data curve. Problems such as, the model used, the number of transitions employed, and nature of the data, can all define the difference between a model with a good fit or a model that marginally describes the calorimetric data. Unlike the UV-melts however, there are less measurement artifacts, such as irregular transitions, that increase the difficulty of obtaining good van't Hoff enthalpy fits. Therefore, the van't Hoff enthalpy fits obtained from calorimetric data offer a higher level of accuracy than is possible with UV-melting techniques. In addition, because the modeled data is based on calorimetric data any error arising as the result of different experimental methods is eliminated. Therefore, accurate comparisons can be made between the modeled and calorimetrically derived data. This is not to say that van't Hoff data derived from other sources is unreliable. The literature is full of many reports where comparisons are made from equilibrium-type van't Hoff measurement and calorimetry. However, if such comparisons are to be made it is better to use data derived from the same source rather than from different methods that are subject to the differences of inherent experimental bias.

Modeling of calorimetric data also provides information with respect to the nature of the transition (whether the all-or-none model is applicable) and the size of the cooperative unit.[152] According to Breslauer, those transitions with significant intermediate states will broaden the integral of the van't Hoff equation (Equation 26 , also shown below) rearranged as an equilibrium expression (Equation 46).

$$d(\ln K)/dT = \Delta H/RT^2$$

Equation 46: $\Delta H_{vh} = (2+2n)RT^2(d\alpha/dT)$

This broadening will lead to a reduced value for the van't Hoff enthalpy relative to the calorimetric enthalpy. Therefore, comparison of the magnitudes of the van't Hoff and calorimetric enthalpies reveal the closeness to an ideal all-or-nothing transition. If the ΔHvh is lower than the ΔHcal, then the transition most likely involves a significant population of intermediate states – the transition is not two-state. If the ΔHvh is equivalent to the ΔHcal, then the transition occurs in a true all-or-none fashion[36]. In addition, the ratio of the ΔHvh to the ΔHcal provides a measure of the fraction of the structure that melts as a cooperative unit. Segments of DNA that are at the cooperative length, will generally, behave as a single unit with near two-state transitions. Thus, the ratio of ΔHvh/ΔHcal gives the approximate length of the cooperative unit in terms of the length of the oligonucleotide under consideration. For example, a 75-bp duplex with a ΔHvh/ΔHcal ratio of 0.5 suggests that approximately one-half of the duplex, or 37-bp, is behaving as a single unit.[153,154]

3.7.4.2 The Experimental Procedures and Models Applied

For all data sets, the calorimetric data was deconvoluted with "CpCalc" software from Calorimetric Sciences Inc. For all data several models, either the general, two-state or "Two-State Scaled" models, were attempted before selection of the final model. Within each model, several different transitional modes were also tested before selection of the final fit. The model and transitional mode that gave the best overall fit, as determined by the value of X^2 was selected for further consideration. Typically, once the model was selected, the data was sequentially modeled with a single, double, and triple transitions until the chi-squared test reached a minima. In addition to the chi-squared test an additional, albeit judgmental, criteria was placed on all data sets. The modeled description of calorimetric data had to make "real-world" sense. For example, fits of multiple melting transitions containing improbable transition points, such as melting transitions at –150 °C or 200 °C, were rejected regardless of the statistical chi-squared fit.

It was determined after numerous trial-and-error attempts that the model best suited for this data set was the "Two-State Scaled" model. Therefore, for all oligomers the excess heat capacity of the calorimetric data was imported into a deconvolution module, and the data modeled with a "Two-State Scaled" model. With this model the transition enthalpy was determined by summing the adjusted enthalpy of the individual transitions. In this equation, Cp^m is the molar heat capacity, ΔH is the enthalpy, T is the temperature in °K, and Aw is a scaling factor used to adjust the scale of the individual enthalpy curve fits (Equation 47).

Equation 47: $\mathbf{Cp^m = (\Delta H/dT)\, Aw}$

This equation is derived from the two-state transition model, that assumes the data has no ΔCp, and that total of the excess heat capacity measurement is the sum of "n": independent transitions.[155] The transition heat capacity is given by the following equation (Equation 48), where the enthalpy for each transition, dH(T) is dependent upon the extent of the transition, and the temperature.

Equation 48: $\mathbf{Cp(T) = dH(T)/dT}$

The enthalpy for each transition is a fraction of the total enthalpy ΔH, which is a constant, according to the following equation (Equation 49). For this relationship, H(T) is the enthalpy of the transition at temperature T, α is the fraction of unfolded molecules, and ΔH is the total enthalpy.

Equation 49: $\mathbf{H(T) = \alpha(T)\Delta H}$

Literally translated, this equation says "the enthalpy of the transition (a function of temperature) is equal to the fraction of molecules unfolded (also a function of the temperature), times the total enthalpy. The fraction of denatured or unfolded molecules as a function of temperature is given by the following equation (Equation 50) where α(T) is the fraction unfolded at temperature T, and K(T) is the equilibrium constant at temperature T.

Equation 50: $\mathbf{\alpha(T) = K(T)/(1+K(T))}$

To translate the equilibrium constant into experimental observables, the following relationship is applied (Equation 51) where K(T) is the equilibrium constant at temperature T, T_m is the transition melting temperature, R is the universal gas constant, ΔH is the total enthalpy.

Equation 51: $K(T) = \exp(-(\Delta H - T(\Delta H/T_m))/RT)$

3.8 Kinetic Characterization

3.8.1 Kinetics as a Function of T (Arrhenius plots, Eyring plots)

3.8.1.1 Essential Principles, and the Arrhenius Equation, Eyring Equation

The Arrhenius equation relates the observed rate constant of a reaction under observation to the activation energy with the following relationship (Equation 52), where E_a is the activation energy in kJ/mol, R is the universal gas constant, k_{obs} is the measured rate constant, and A is the collision frequency factor.

Equation 52: $\ln(k_{obs}) = \ln A - E_a/RT$

A plot of $\ln(k_{obs})$ vs. 1/T yields a line with a slope equal to –Ea/R and intercept of the collision frequency factor. Because this information, while useful, does not describe the enthalpy or entropy of activation, it is necessary to redefine the data plot. To obtain thermodynamic information with respect to the transition state parameters, the kinetic data can be plotted using the Eyring equation[156], (Equation 53)

Equation 53: $k_{obs} = (k_b T/h)\, e^{-\Delta G/RT}$

where k_{obs} is the observed rate constant, k_b is Boltzman's constant, T is the absolute temperature, h is Planck's constant, ΔG is the activation free energy, R is the universal gas constant.[156] The Eyring equation relates thermodynamic parameters to the transition state by relating ΔG, the free energy to the activation energy term, Ea, of the Arrhenius equation. The Gibbs free energy relationship (Equation 54),

Equation 54: $\Delta G = \Delta H - T\Delta S$

can be substituted for ΔG in the Eyring equation (),

Equation 55: $k_{obs} = (k_b T/h)\, e^{-\Delta H/RT + \Delta S/R}$

which can be rearranged to produce the following relationship for plotting kinetics data (Equation 56).

Equation 56: $\ln(k_{obs}/T) = -(k_b/h)\,\Delta H/RT + \Delta S/R$

A plot of $\ln(k_{obs}/T)$ versus 1/T describes a line with slope of $-\Delta H k_b/Rh$ and intercept of ΔS/R. The transition enthalpy and entropy terms can also be calculated from the following relationship that avoids the use of graphical analysis of the data (Equation 57).[157]

Equation 57: $\Delta H_{ts} = E_a - RT$

3.8.1.2 Sample Preparation and Procedures

All DNA solutions were prepared from purified, characterized and concentrated DNA stock solutions to achieve a final concentration between 0.5 μM and 2 μM DNA in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, 100 mM NaCl pH 7.0. Typically, the DNA stock solution was a former DSC-sample, at high concentration in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, 100 mM NaCl pH 7.0 buffer. Preparation of kinetics samples from the former DSC samples was achieved by diluting an aliquot of the DNA with 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, 100 mM NaCl pH 7.0 buffer. It was found previously, that kinetics measurements required DNA concentrations sufficient to obtain absorbance signals greater than 0.2 A_{260} (to enable accurate

measurement of the change in absorbance) with rates that were neither too fast to measure accurately nor too slow to measure in a reasonable time period (partially dependent on the starting DNA concentration). While the strand concentration of strands within a triplet repeat series, such as CAG4, CAG10, and CAG25, varied between 0.3 and 0.5 A_{260}, the concentration of complementary strands, such as CAG4 and CTG4, were fixed at equimolar concentrations.[158] The samples were heated from 90 – 100 °C for 15 minutes then slowly cooled to room temperature over 60 minutes. The sample was transferred to 4 °C storage and allowed to equilibrate for a minimum of 48 hours. Prior to kinetics measurement two 4 ml sample equilibration cuvettes (called the equilibration cuvette) and six 0.5 ml cuvettes (called the kinetics measurement cuvette) were thermally equilibrated at the start temperature for 15 minutes. The equilibration cuvettes were placed in positions 8 and 10 of the multicell holder, the kinetics cells were placed in positions 1 – 6 of the multicell holder. Once thermal equilibration was achieved, approximately 3 ml of complementary DNA solutions were transferred from storage at 4 °C to separate 4 ml equilibration cuvettes inside of the multicell holder. The solutions were then equilibrated at the start temperature for a minimum of 20 minutes prior to further execution of kinetics protocols. After this time, 500 µl of one strand ("Watson" strand) was transferred to the kinetics measurement cell, the kinetics cell flow cap (KCFC) was inserted in the top of the cuvette (Figure 13), and the sample was allowed to equilibrate for an additional 5 minutes. During this time, the absorbance at 260 nm was noted for subsequent calculation of the t = 0 DNA concentration. Data collection was initiated prior to reaction initiation to ensure that the spectrophotometer was actively collecting data, and to ensure that the true t = 0 kinetics data were collected. To initiate

the reaction 500 µl of complementary DNA solution drawn from the equilibration cuvette and immediately injected into the kinetics measurement cell through the upper filling line of the KFCC (Figure 13). Mixing within the cell was rapidly accomplished by "pumping" the pipettor up and down three to four times over 5 seconds. Progress of the reaction was monitored as a function of absorbance decay versus time. In total, a minimum of four data points were collected for each temperature and for each replicate the identity of the "Watson" strand was alternated between the two oligonucleotides to minimize any effects the order of sample addition may have had on the experiment.

Because of limitations within the Varian "Kinetics" software all data files were manually processed to remove the portion of the data collection that included the mixing portion of experiment. Prior to export, the kinetic t = 0 time point of the reaction data set was noted and then all data files were converted into Microsoft Excel-readable '.CSV files with the Varian "Kinetics" software (Figure 35). An Excel macro was prepared to expedite removal of all mixing data prior to the previously noted t = 0 time point, and all subsequent kinetics data were offset to the adjusted zero point. The Excel-corrected data file was imported back into Varian "kinetics" software and the data sets fit with a second order rate equation using the DNA concentration and the integration start/stop points as the input parameters (Figure 36). The second order rate equation (Equation 58),

Equation 58: $$A_t = ((A_0 - A_\infty) + A_\infty) / ((X_0 k_{obs} t) + 1)$$

where A_t is the absorbance at time t, A_0 is the absorbance at time t = 0, A_∞ is the absorbance at time t = infinity, k_{obs} is the observed rate constant, t is time in minutes, and X_0 is the initial concentration of one component.

Figure 35: The raw data for a HP to DUP association.

The raw data including the premixing absorbance measurement, the mixing portion, and the association reaction are shown for the association of CAG15 with CTG15 in 10 mM sodium phosphate, 0.1 mM EDTA, 100 mM sodium chloride, pH 7.0.

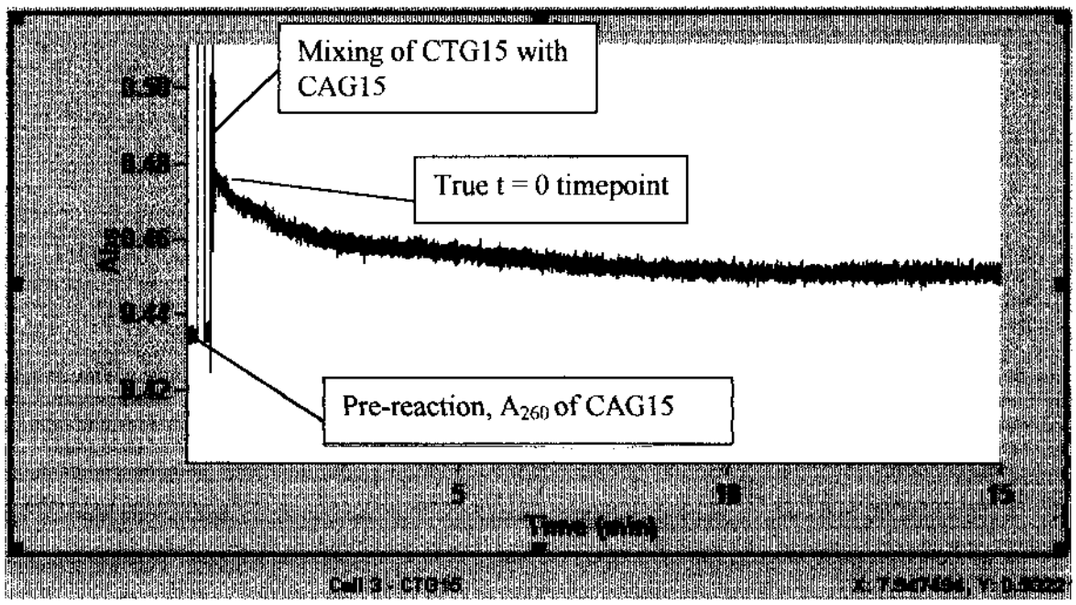
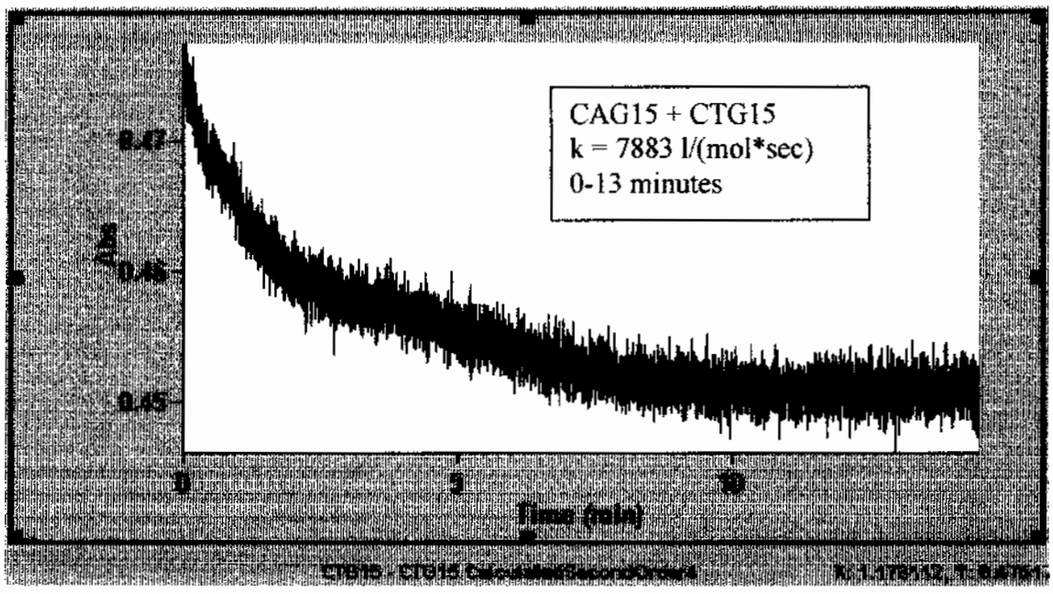


Figure 36: Excell-corrected kinetics data with a 2nd order fit.

The Excel-corrected data (blue), with the Varian "Kinetics" software 2nd order fit (red) are shown for the association of CAG15 with CTG15 in 10 mM sodium phosphate, 0.1 mM EDTA, 100 mM sodium chloride, pH 7.0. All data were corrected to remove the mixing portion of the experiment, and to off-set all data points to the t = 0 of the association reaction.



The observed rate constants were entered onto another Excel spreadsheet for collation and, ultimately conversion into the X and Y coordinates for the type of experimental plot information sought. For Arrhenius plots, the observed rate constants were converted from k_{obs} in units of l/(mol min) to k_{obs} in units of l/(mol sec), then finally into "$\ln k_{obs}$" and the temperature data converted into the $1/^{o}K$. The slope of this plot gave $-E_a/R$ or the activation energy of the transition, and the intercept gave the collision frequency factor. For Eyring plots, the data were converted from k_{obs} to $\ln(k_{obs}/T)$ vs $1/^{o}K$. The slope of this plot gave the transition enthalpy as $-\Delta H/R$ and the intercept gave the transition entropy.

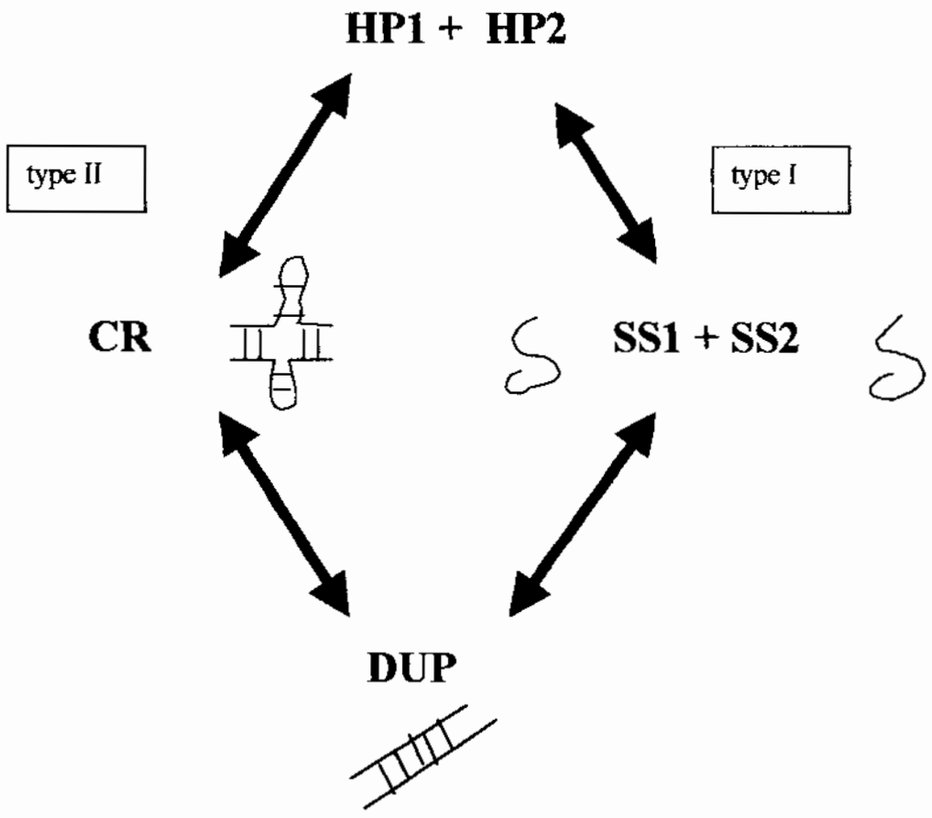
3.8.2 Kinetics as a Function of Salt Concentration

3.8.2.1 Essential Principles

The behavior of the HP to DUP renaturation kinetics as function of sodium concentration can yield information about the mechanism through which TR-DNA strands anneal. Such information is relevant to the overall disease mechanism because it will define the renaturation process and possibly explain the underlying reasons why certain sequences expand more frequently, or why long strands are more likely to expand. The theoretical framework for such experiments comes from the work of Azivonas and Kearns, who investigated the mechanism through which DNA hairpins anneal to duplex DNA forms.[38] The formation of duplex DNA from DNA hairpins containing complementary sequences can proceed via two pathways (Figure 37). One pathway, dubbed the single strand (SS) pathway, involves complete hairpin strand denaturation followed by nucleation to the complementary sequence, and the formation

Figure 37: The hairpin (HP) to duplex (DUP) interconversion pathway of Azivonas.

The type I mechanism is characterized by a single-stranded intermediate and will be destabilized by salt. Increasing the salt concentration will increase the rate of the forward HP to DUP association process. Because many hydrogen bonds are broken simultaneously, the energy of activation should be as large or larger than the free energy of HP stability. The type II mechanism is characterized by a cruciform intermediate and, because this intermediate resembles duplex DNA, will be stabilized by salt. Increasing salt concentration will increase the rate of the reverse CR to HP dissociation process. This should decrease the rate of the forward HP to DUP association. Because only a few hydrogen bonds are broken at any given time, the energy of activation should be smaller than the free energy of hairpin stability.



of the double stranded duplex. The other pathway, dubbed the cruciform (CR) pathway, involves the partial melting of the hairpin DNA stands, and nucleation at the ends to form a cruciform intermediate. The hairpin arms of the cruciform intermediate then melts and associates with the complementary sequence forming a duplex DNA strand. Because the two pathways go through different intermediates, either the cruciform or single-strand species, there are unique properties associated with the transition states specific to each pathway. The SS pathway entails the breaking of many hydrogen bonds as the hairpins are denatured to the single strand state. Therefore, the activation energies for such a pathway should be as high or higher than the free energies required to melt the HP strands (termed the transition free energy or the free energy of stability). In addition, because the single stranded form is destabilized by sodium ions and the duplex form stabilized by sodium ions, the SS pathway should be activated with increasing sodium ion concentration. The observed HP to DUP rate constants should therefore increase as the concentration of sodium increases. On the other hand, the CR pathway entails the breaking of a few H-bonds prior to nucleation and association to the complementary DNA strand. This pathway should be characterized by low activation energies, because of the low number of bonds broken at any given time. In addition, because the intermediate of the pathway resembles duplex-DNA, which is stabilized by sodium ions, the forward reaction will be less favored as sodium ion concentration increases – there is no underlying "incentive" for the reaction to proceed if the intermediate is stabilized. This stabilization allows the intermediate to easily decay back to meta-stable hairpin resulting in an increase in the rate constant of the reverse reaction (CR decaying to HP). Because increasing the salt concentration favors the reverse CR to HP reaction, the

observable forward rate constant (HP to DUP) will either decrease or remain unaffected as a function of increasing sodium concentration.

Application of this technique requires exhaustive measurement of individual activation energy parameters as a function of each salt concentration.[159] Due to the size of such a study, an abbreviated plan was undertaken to determine if the mechanistic trends could be described at a single temperature (37 °C) as a function of sodium concentration.

3.8.2.2 Sample Preparation and Procedures for Kinetics vs. Salt

All DNA solutions for the purposes of kinetics measurements were prepared from purified, characterized, and concentrated DNA stock solutions in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH 7.0 to a final concentration between 0.2 μM and 2.0 μM DNA in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH 7.0 with salt concentrations from 15 mM to 500 mM. It was found previously, that kinetics measurements required DNA concentrations sufficient to obtain good absorbance signals (greater than 0.2 A_{260} to enable accurate measurement of the change in absorbance) with rates that were neither too fast to measure accurately nor too slow to measure in a reasonable time period (partially dependent on the starting DNA concentration). While the concentration of strands within a triplet repeat series, such as CAG4, CAG10, and CAG25, varied between 0.3 and 0.5 A_{260}, the concentration of complementary strands, such as CAG4 and CTG4, were fixed at equimolar concentrations. The samples were heated from 90 – 100 °C for 15 minutes then slowly cooled to room temperature over 60 minutes. The sample was transferred to 4 °C storage and allowed to equilibrate for a minimum of 48 hours. Prior to kinetics measurement two 4 ml sample equilibration

cuvettes (called the equilibration cuvette) and six 0.5 ml cuvettes (called the kinetics measurement cuvette) were thermally equilibrated at the start temperature for 15 minutes. The equilibration cuvettes were placed in positions 8 and 10 of the multicell holder, the kinetics cells were placed in positions 1 – 6 of the multicell holder. Once thermal equilibration was achieved, approximately 3 ml of complementary DNA solutions were transferred from storage at 4 °C to separate 4 ml equilibration cuvettes inside of the multicell holder. The solutions were then equilibrated at the start temperature for a minimum of 20 minutes prior to further execution of kinetics protocols. After this time, 500 µl of one strand ("Watson" strand) was transferred to the kinetics measurement cell and the kinetics flow cell cap (KFCC) was inserted in the top of the cuvette. The sample was allowed to equilibrate for an additional 5 minutes. During this time, the absorbance at 260 nm was noted for subsequent calculation of the t = 0 DNA concentration. Data collection was initiated prior to reaction initiation to ensure that the spectrophotometer was actively collecting data, and to ensure that the true t = 0 kinetics data were collected. To initiate the reaction 500 µl of complementary DNA solution at an equimolar concentration was drawn from the equilibration cuvette and immediately injected into the kinetics measurement cell (Figure 13). Mixing within the cell was rapidly accomplished by "pumping" the pipettor up and down three to four times over 5 seconds. Progress of the reaction was monitored as a function of absorbance decay versus time. In total, a minimum of four data points were collected for each temperature and for each replicate the identity of the "Watson" strand was alternated between the two oligonucleotides to minimize any effects the order of sample addition may have had on the experiment.

Because of limitations within the Varian "Kinetics" software all data files were manually processed, with Excel software, prior to determination of the observed rate constants. Prior to export into Microsoft Excel, the kinetic t_0 time point of each measurement, defined as the point in time after manual mixing was complete, was noted and then all data files were converted into Excel-readable "CSV" files with the Varian "Kinetics" software. An Excel macro was prepared to expedite removal of all mixing data prior to the t_0 time point, and all subsequent kinetic data were offset to the adjusted zero point by subtraction. The Excel corrected data file was imported, as a ".CSV" file, into Varian "kinetics" software and the data sets fit with a second order rate equation (Equation 59)

Equation 59: $$A_t = ((A_0 - A_\infty) + A_\infty) / ((X_0 k_{obs} t) + 1)$$

where A_t is the absorbance at time t, A_0 is the absorbance at time t = 0, A_∞ is the absorbance at time t = infinity, k_{obs} is the observed rate constant, t is time in minutes, and X_0 is the initial concentration of one component. The DNA concentration and the integration start/stop points were the user-specified input parameters.

4 RESULTS

4.1 Sample Characterization

4.1.1 The Oligonucleotides Under Study

The triplet repeat sequences, (CNG)n where N = A, C, G, or T and n = 4, 5, 10, 15, or 25 were examined from a variety of perspectives so that the biophysical behavior of these sequences could be fully characterized. The sequences were synthesized with β-cyanoethylphosphoramidite synthesis procedures in high yield and efficiency, as shown by the trityl assay (Figure 22,A). The oligonucleotides were purified with either RPHPLC or PolyPak II procedures, and the final purities determined by a combination of RPHPLC (Figure 24,C), or AXHPLC with UREA-PAGE (Figure 27, A). For all experiments, the oligomers were purified to either single-band purity by UREA-PAGE analysis, or they were rejected.

4.1.2 Comparison of Purification Procedures, RPHPLC to PolyPakII

Purification of the oligomers by PolyPak II (PP II) reversed-phase cartridge columns introduced a new method for DNA purification. Like all new methods the equivalency, or superiority if applicable, between the new technology (PP II) and the old (RPHPLC) had to be demonstrated before the new technology could be adopted. Validation of the PPII technology, however, was not undertaken in a single experiment, but rather as the collected data of several experiments where full comparisons could be drawn. A comparison of the purities and yields obtained with the PolyPak II cartridge columns relative to that obtained with RPHPLC had shown subtle differences in purity, yet marked differences in yield. Urea-PAGE results for the purification of G6T4G6 had

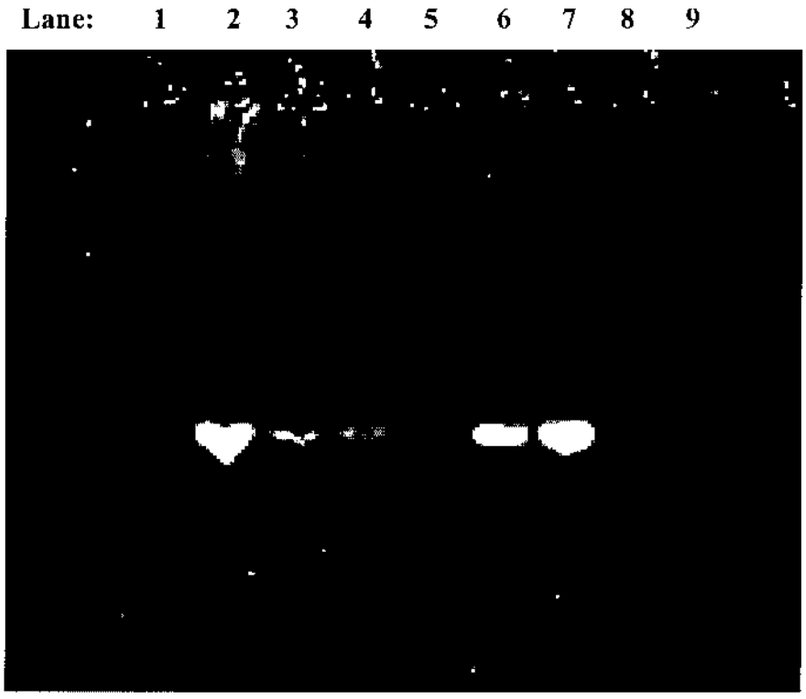
shown that the purity between the two methods were similar (Figure 38). The yield of the PPII-purified oligomer CCG10 (CCG10.PP11.101501) was approximately 961μg of purified DNA in 1.0 ml, or approximately 100 μM DNA in 1.0 ml. In contrast, an RPHPLC purified CCG10 oligomer resulted in an overall yield of 692 μg of purifed DNA in 1.0 ml, or approximately 72 μM DNA in 1 ml (oligomer CCG10.071001.01). As has been demonstrated for these oligomers, the PolyPak II protocols can result in pure DNA with high yield. However, in the course of this exercise, a few limitations related to the PP II columns were found. First, because the PP II columns have lower resolving capabilities than RPHPLC, they were limited to the purification of oligomers from relatively pure synthesis reactions. Second, for reasons that remain undefined, the yields from column to column can be variable even with identical oligomers. Third the flow rates were found to be critical to success – slow flow rates produced consistently higher yields than did fast flow rates. While the slower flow rates were found to limit the purification through-put, the overall productivity with PP II columns were still higher than that achievable with RPHPLC.

4.1.3 Comparison of Dialyzed vs. PolPakII-washed Oligomers

Currently, the typical practice within the DSC literature entails dialysis of the sample into the experimentally relevant buffer prior to calorimetric analysis. The measurement of the DNA calorimetric parameters, however, required quantities of DNA that either had to come from a high-yield purification at the 1-μmole scale, or from a series of medium-yield purifications with DNA syntheses at higher scales. It was found that dialysis of the triplet repeat oligomers could result in losses of up to 50% for 12-mers and up to 20% for 30-mers or 45-mers. These losses were inconsistent from oligomer to oligomer even

Figure 38: Comparison of purity between PP II and RPHPLC purified oligomers.

A purity comparison by Urea-PAGE of the oligomer G6T4G6 purified by either PolyPak II reversed-phase purification cartridges, or RPHPLC. Lanes 1-4, the oligomer at various stages of the PolyPak II purification process: lane 1, crude oligomer; lane 2 column flow through; lanes 3,4 acetonitrile eluates. Lanes 5-9 are RPHPLC fractions of the detritylated DNA peak (the secondary purification), beginning, from left to right fractions 3 through 8. Fractions 5 and 6 of this chromatogram represent the main peak fractions, and are also display the highest level of purity within the chromatogram.



within the same sequence and length. The loss incurred by dialysis in combination with the loss incurred by the RPHPLC purification process were too great to accomplish the goals of this project. Therefore, an alternative method was devised to purify oligomers with greater yield, yet still maintain the high buffer and sample purity that can be accomplished through dialysis. The synthesizer optimization protocols improved synthesis efficiencies to the extent that little purification of the crude oligomer was required. In addition, the PolyPak II purification protocols improved the final oligomer yield while at the same time provided a basis for desalting the crude oligomer. The PolyPak II purification procedures however, were novel, and dependent upon the assumption that the exchange of salts or unwanted organic species was a fast, kinetically driven process. If this assumption was incorrect, then incomplete salt removal would occur and the buffer composition of the prepared sample would differ from that of the reference buffer. The greatest danger in preparing samples with the PPII method is that the DSC thermograms will be different for a PPII-washed oligomer than the same oligomer were it dialyzed prior to analysis. Therefore, a comparison of representative oligomers was undertaken to demonstrate the equivalence, in a calorimetric sense, of the two methods for buffer exchange. A comparison of oligomers shows that with respect to the T_m (Table 15), and the nature of the thermograms (Figure 39) show little, if any, difference in the thermodynamic parameters. It was concluded that desalting by the PolyPak II purification procedures produced oligomers equivalent to those produced by dialysis.

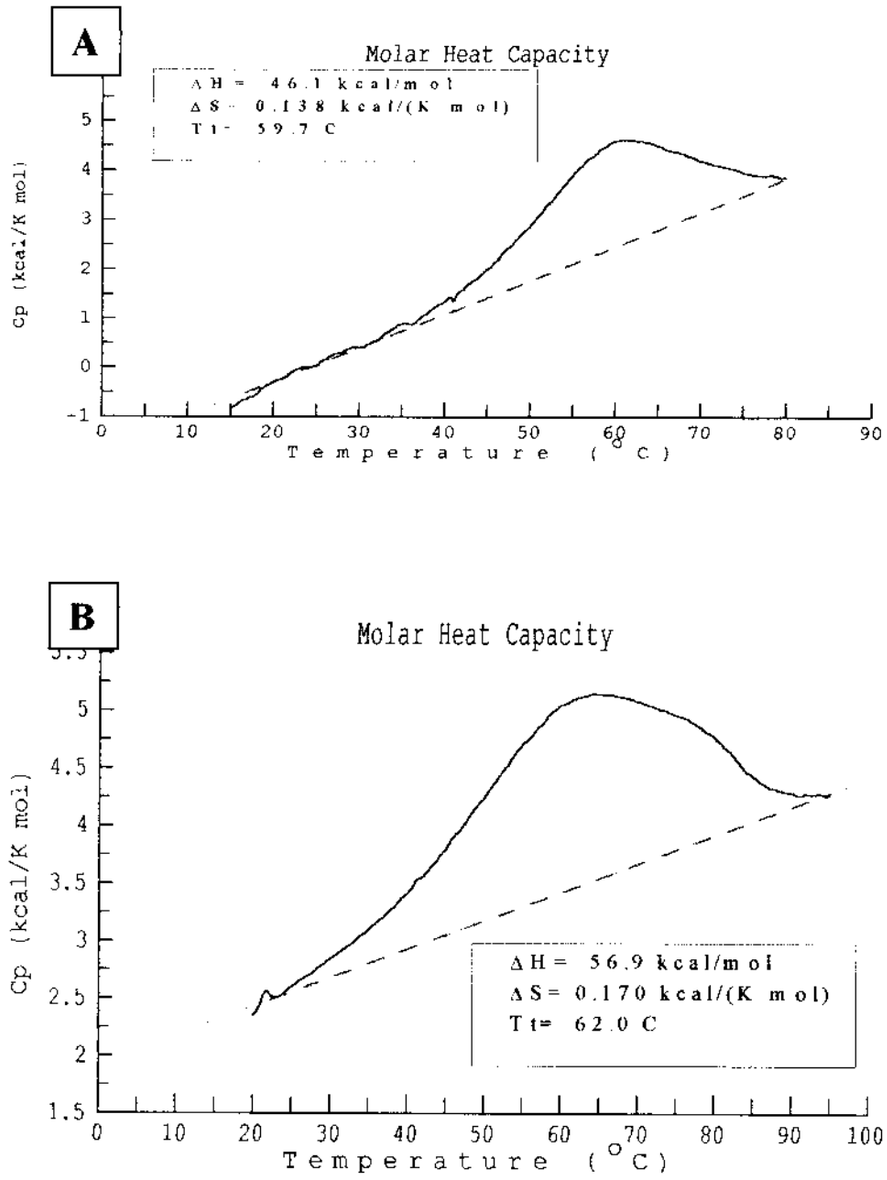
Table 15: Comparison of melting temperatures, PPII vs. dialysis.

A comparison of the melting temperatures, derived by DSC, between DNA samples that were dialyzed or extensively washed/desalted by PolyPak II RP-cartridge columns shows variations that are within the experimental error for these measurements.

	T_m	T_m
OLIGOMER	PolyPak II	Dialysis
CAG4	61	64
CAG10	62	61
CAG15	62	60
CCG4	53	52
CCG10	50	49
CCG15	51	49
CGG10	82	83
CGG15	85	85
CTG4	53	53
CTG10	65	65
CTG15	66	64

Figure 39: A comparison of DSC thermograms between PPII and dialyzed samples.

A. The DSC thermogram of CAG10 (CAG10.032201.01.CPC) purified by PolyPakII procedures and dialyzed with traditional dialysis techniques. **B.** The DSC thermogram of CAG10 (CAG10.122001.01.CPC) purified by PolyPakII procedures with on-column washing as described in "Materials and Methods" to accomplish salt removal.



4.1.4 Comparison of the [DNA] Differences based on SEC or AEC

The differences in the extinction coefficients derived by either the summed extinction coefficient (SEC) or averaged extinction coefficient (AEC) method made it prudent to compare the error in DNA concentration between the two methods. The results (Table 16) show that the AEC-method over-estimated the μmolar DNA concentration by 5% for the CCG, CTG series of solitary oligomers, 5% for all duplex DNA concentrations, approximately 15% for the CGG series of oligomers, and 35% for the CAG series of oligomers. The estimated concentration differences for the CAG and CGG oligomers were above the limits of precision for all assays mentioned. Because of the observed differences, the SEC was used for all assays where the precision of the DNA concentration needed to be within 5% or better, and whenever cross-comparisons between assays were made. Concentrations determined by the AEC method were reserved for routine, non-critical measurements where high sample processing was required, or whenever a comparison of the change in DNA concentration within an experiment was performed (for example, measurement of the pre/post dialysis DNA concentration).

Table 16: Comparison of DNA concentrations.

The AEC method overestimates the [DNA] from 1 – 35% for all sequences compared. The SEC-derived [DNA] was used for all assays where [DNA] had to be known to a high degree of certainty.

		[DNA] uM	[DNA] uM	
OLIGOMER	A260	S.E.C	A.E.C.	Difference
CAG4	1	7.33	9.90	0.35
CAG5	1	5.86	7.96	0.36
CAG10	1	2.93	3.98	0.36
CAG15	1	1.95	2.66	0.36
CAG25	1	1.17	1.60	0.36
CCG4	1	9.54	9.65	0.01
CCG5	1	7.63	7.70	0.01
CCG10	1	3.82	3.85	0.01
CCG15	1	2.54	2.57	0.01
CCG25	1	1.52	1.54	0.01
CGG4	1	8.03	9.25	0.15
CGG5	1	6.43	7.41	0.15
CGG10	1	3.21	3.70	0.15
CGG15	1	2.14	2.47	0.15
CGG25	1	1.28	1.48	0.16
CTG4	1	9.09	9.50	0.04
CTG5	1	7.27	7.58	0.04
CTG10	1	3.63	3.79	0.04
CTG15	1	2.42	2.53	0.04
CTG25	1	1.45	1.52	0.05
CAG4/CTG4	1	6.31	6.54	0.04
CAG5/CTG5	1	5.05	5.24	0.04
CAG10/CTG10	1	2.52	2.62	0.04
CAG15/CTG15	1	1.68	1.75	0.04
CAG25/CTG25	1	1.01	1.05	0.04
CCG4/CGG4	1	6.31	6.37	0.01
CCG5/CGG5	1	5.05	5.10	0.01
CCG10/CGG10	1	2.52	2.55	0.01
CCG15/CGG15	1	1.68	1.70	0.01
CCG25/CGG25	1	1.01	1.02	0.01

4.2 Structural Studies

4.2.1 DNA Structure by Circular Dichroism Spectroscopy

To date, the structures formed by triplet repeat DNA sequences remains defined for only a limited data set. Hence the structures formed by all of the triplet repeat sequences was characterized in a systematic fashion to not only complete the data base, but to also define the end-states for subsequent biophysical studies. The structures formed by triplet repeat DNA sequences were characterized by a variety of orthogonal techniques. Included within the many techniques utilized was circular dichroism spectroscopy because this technique is uniquely suited for the characterization of DNA structure and structural transitions. Because of this fact, the circular dichroism spectrum for all (CNG)n oligonucleotides where N = A, C, G, or T and n = 4, 10, 15, or 25 (12, 30, 45, or 75-mer respectively) was measured at 25 °C, in 5 mM monobasic sodium phosphate, 5 mM dibasic sodium phosphate, 0.1 mM EDTA, 100 mM sodium chloride, pH 7.0, from 320 nm to 210 nm.

In general, the CD spectra of the CAG, CCG, and CTG series of oligomers were similar to each other and, with the exception of the CGG oligomers, were similar to the CD spectral signature of B-form DNA (Figure 40, and Figure 41). Regular B-form DNA has a peak centered in the region of 275 - 280 nm, two cross-over points centered at 215 and 265 nm, and a trough centered in the region of 240 - 255 nm.[35] The CD spectra of the (CAG)n oligomers, adjusted for oligomer concentration, presented an asymmetric peak centered at 275 nm with a shoulder centered at approximately 290 nm (Figure 40, A). For all lengths studied, there were two isoelliptic crossover points, centered at 260 nm

Figure 40: The CD spectra of the solitary (CAG)n and (CCG)n oligonnucleotides.

The CD spectra for the (CNG)n triplet repeat sequences at 25 °C in 5 mM monobasic sodium phosphate, 5 mM dibasic sodium phosphate, 0.1 mM EDTA, 100 mM sodium chloride, pH 7.0. The CD spectra are similar to that of B-form DNA structures. All CD spectra have been adjusted for oligomer concentration. **A**. The CD spectra of the (CAG)n oligomers where n = 4 – 25 (12 – 75 nucleotides in length). **B**. The CD spectra of the (CCG)n oligomers where n = 4 – 25 (12 – 75 nucleotides in length).

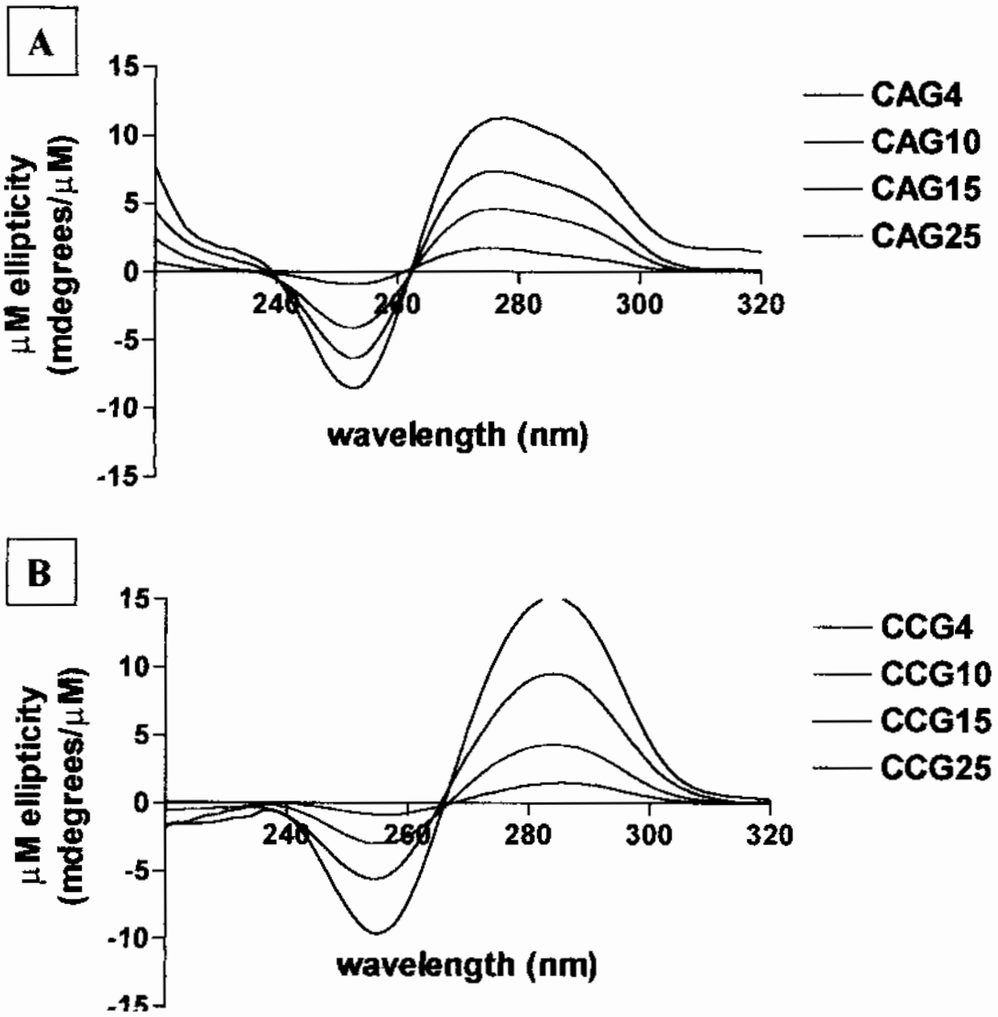
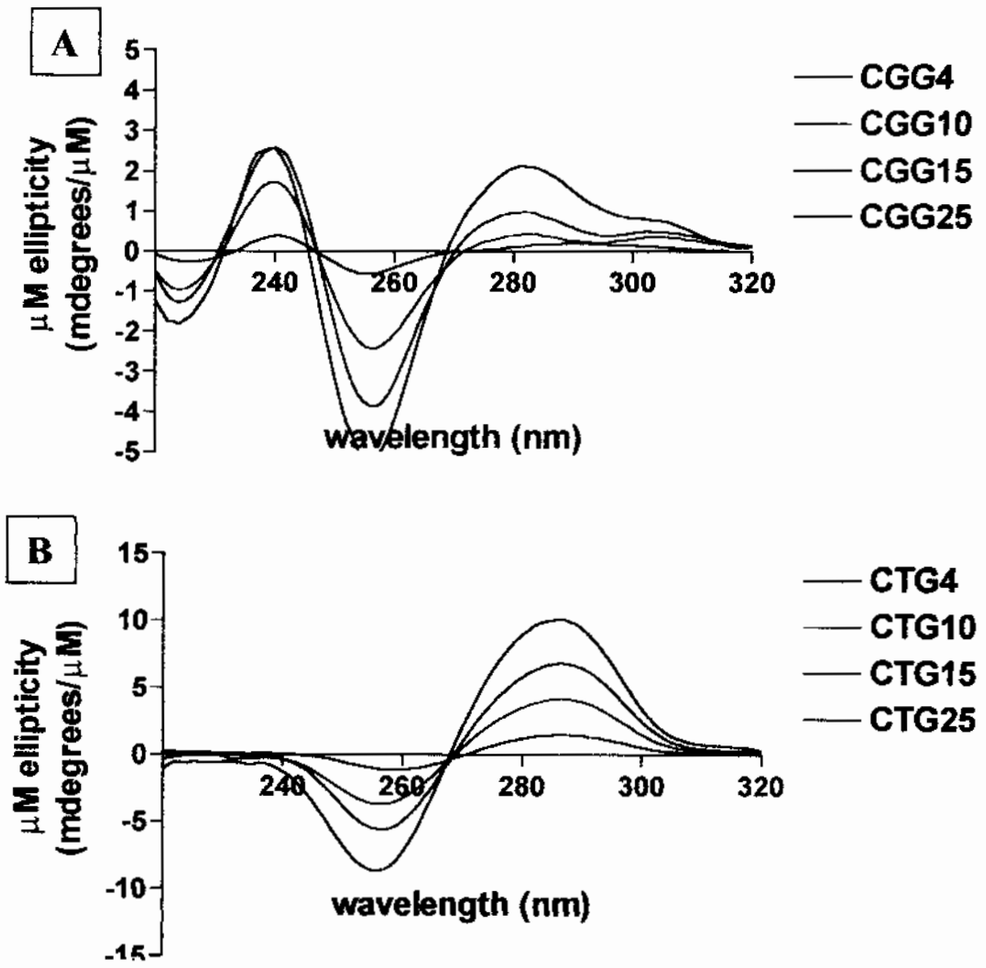


Figure 41: The CD spectra of the solitary (CGG)n and (CTG)n oligonnucleotides.

The CD spectra for the (CNG)n triplet repeat sequences at 25 °C in 5 mM monobasic sodium phosphate, 5 mM dibasic sodium phosphate, 0.1 mM EDTA, 100 mM sodium chloride, pH 7.0. With the exception of the CGG series of oligomers, the CD spectra are similar to that of B-form DNA structures. All CD spectra have been adjusted for oligomer concentration. **A**. The CD spectra of the (CGG)n oligomers where n = 4 – 25 (12 – 75 nucleotides in length). **B**. The CD spectra of the (CTG)n oligomers where n = 4 – 25 (12 – 75 nucleotides in length).



and 238 nm, and a trough at 251 nm. The far UV region of the CD spectra near 220 nm was positive, with a slope that became incrementally higher as a function of oligomer length. The positive slope of the far-UV region suggests that the CAG oligomers form right-handed helices.[160] The height of the peak centered at 275 nm increased as a function of length, indicating base stacking interactions increased as the length of the CAG oligomer increased. Likewise, the depth of the troughs centered at 251 nm became more negative as a function of increasing length, further supporting the observation that the base stacking interactions increased as function of length. The CD structural studies suggest that the structure becomes more stabilized as length increases due to increased base stacking interactions. In addition, the CD spectra suggest that the overall nature of the structures formed by the CAG oligomers does not significantly change as the length is increased from a 12-mer to a 75-mer. Surprising, still is the fact that an oligomer as short as 12 nucleotides long can form a structure as shown by the CD spectra.

The oligomer concentration adjusted CD spectra of the (CCG)n series of oligomers produced a symmetric peak centered at 284 +/- 1 nm, an isoelliptic crossover occurring at 265 nm, with a trough centered at 254 +/- 1 nm (Figure 40, B). The height of the observed peaks initially increased from CCG4, the 12-mer, to CCG15, the 45-mer, but then decreased approximately 5 millidegrees/$[DNA]_{\mu molar}$ for CCG25, the 75-mer. In addition, the same trend was observed for the depth of the trough at 254 nm. The depth of the trough was more negative up to the 45-mer but then became positive for the 75-mer when adjusted for DNA concentration. These results suggest that the base stacking interactions increased as a function of length until a certain length threshold was

achieved. Because the base stacking interactions contribute to the enthalpy of stability, the trend predicts that the enthalpy of the oligomeric stability will increase up to about 45 nucleotides then either decrease or remain unchanged for all oligomers beyond 50 nucleotides long. Surprisingly, as was observed for the CAG series, oligomers as short as 12 nucleotides long formed stable structures. Overall, the nature of the structure formed by CCG oligomers did not significantly change over the range of lengths studied, as shown by the overall similarity between spectra and identical crossover points.

The CD spectra of the (CTG)n oligomers, adjusted for oligomer concentration, produced a symmetric peak centered at 286 nm, and an isoelliptic crossover at 268 nm (Figure 41, B). The CD spectra had also shown a trough centered initially at 259 nm for the 12-mer, which became noticeably blue-shifted to 255 nm as the length of the oligonucleotide increased towards the 75-mer. The blue-shift suggests that the nucleotides, most likely the mismatched nucleotides, are becoming more hydrated or exposed to solvent as a function of length. Hydration suggests a decrease in the entropy of water due to the formation of clathrate structures around the exposed hydrophobic surfaces, while at the same time a possible increase in the entropy of the nucleotide due to an increase in the conformational entropy. This also suggests that the sequence undergoes subtle changes in the orientation of the nucleotides, presumably to achieve an overall stable conformation, as length increases. The observed base stacking increase can be due either to an increase in the overlap of the nucleotides or simply the result of the cumulative effects of stacking interactions summed over the entire molecule. Overall, the increase in the base stacking interactions suggests this sequence should have increased stabilities as a

function of length. Furthermore, the overall nature of the CD spectra, from the similar shape to the identical crossovers point for all strands suggest that the CTG oligomers do not undergo major changes in structure, but rather undergo subtle changes such as reorientation of the nucleotide bases.

In a more general view, the locations of the peak maxima and peak minima (troughs) and the location of the crossover points suggest that the solitary oligonucleotides are forming DNA structures, dubbed "B-like", that are similar to B-form DNA. On the basis of CD spectroscopy it is difficult to determine if these structures are unimolecular hairpins, or mismatched duplexes. Both structures can assume a B-like conformation, and both structures must contain a mismatched base-pair every third nucleotide for stable complementarity of the structure as a whole. The CD spectral data suggest that orientation of the mismatched nucleotides can change as a function of length, perhaps to achieve greater stability, and sequence. The presence of the mismatch within the sequence can also serve to alter the CD spectral signal by either providing a water-accessible hydrophobic surface within the DNA structure, as is the case with exposure of a nucleotide mismatch (blue-shift), or by burying otherwise exposed surfaces, as could happen with reorienting basepairs to bury hydrophobic groups (red-shift). The net effect of such subtle changes to the DNA structure is to alter the classic B-form DNA spectra. For the CAG oligomers, the relative blue shifting in the DNA peaks suggests that the mismatched nucleotides are being reoriented, relative to fully duplexed DNA, towards the aqueous environment. Relative to the CAG oligomers, the CCG and CTG oligomers had peak maxima and minima that were red-shifted. Relative to the 275 nm peak for

(CAG)n, the (CCG)n oligomers had a red-shifted maximum of 285nm and the (CTG)n oligomers had a maximum of 286 nm suggesting that there is less hydration surrounding the either the mismatched nucleotides or the base-paired nucleotides. It is interesting to note that while the troughs of the CCG and CTG oligomers began within the "red" region of the spectrum, they eventually became blue-shifted with increasing length. At 75 nucleotides, the trough is almost at the B-form DNA benchmark. This suggests that the DNA is achieving a more stable structure as the length increases by increased hydration of the mismatched and/or base-paired nucleotide, while at the same time becoming becoming more B-form in nature.

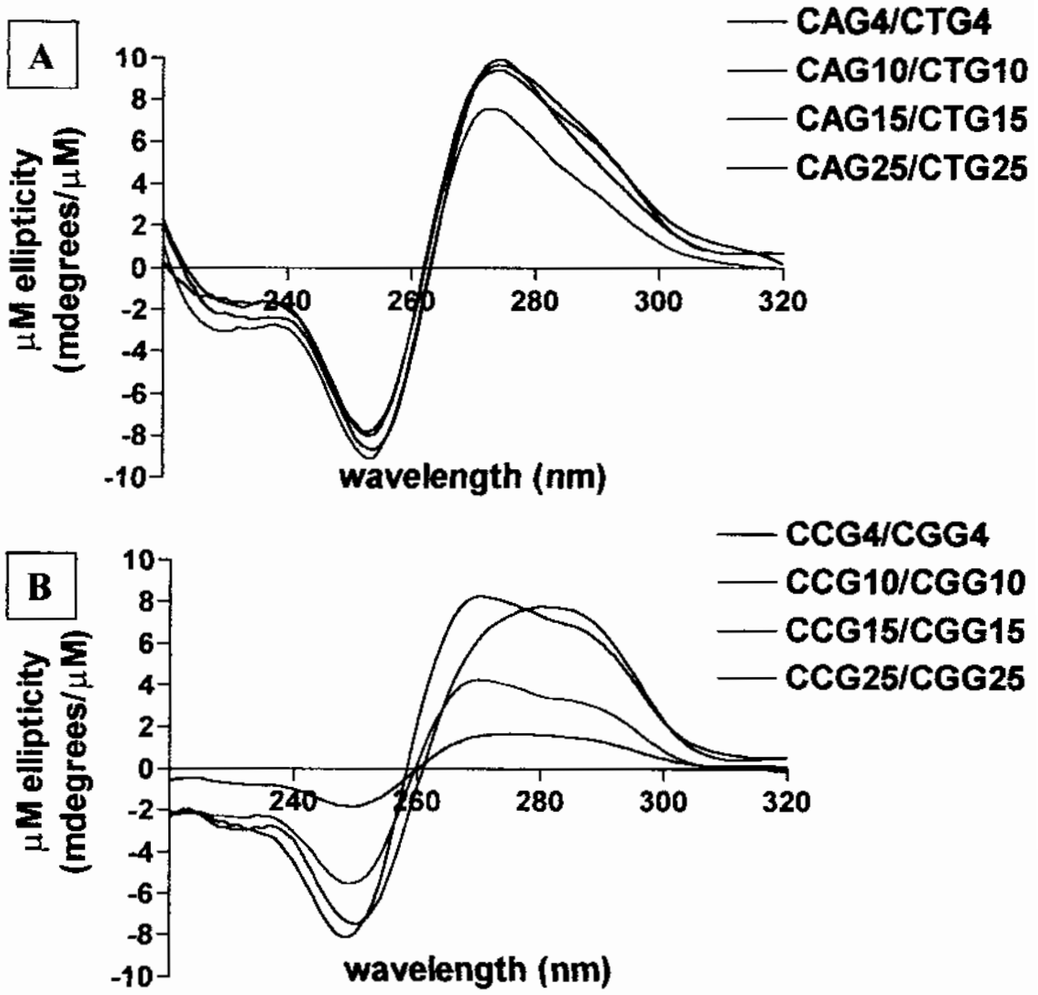
The CD spectra for the (CGG)n oligomers are in a structural and spectral "class of their own" that cannot be readily assigned by referral to a reference spectrum (Figure 41, A). In addition, there was a significant change in structure as the length increased from 12 nucleotides to 30 nucleotides. Beyond 30 nucleotides the structure, whatever it may be, remained stable up to the 75-mer. The (CGG)n oligomers where n = 10 to n = 25 showed several peaks centered at 303 nm, 282 nm, and 240 nm and two troughs, or minima, located at 256 nm and 223 nm. There were two crossover points, with length-dependent isoellipticities located at 268 nm that is isoelliptic for the 30-mer through 75-mer, and at 247 nm that is isoelliptic for the 12-mer through 45-mer. The CD spectrum of the 12-mer differed from that of the longer oligomers, with two peaks centered at 292 nm and 240 nm and two troughs centered at 255 nm, and 225 nm. The overall trends of the spectra suggest that the base stacking interactions, as indicated by the increase in peak height and the increase in trough depth, becomes larger as the length of the oligonucleotide

increases. The increases predict that the stability of the CGG oligomers will incrementally increase as the length of oligomer increases. In addition, the positive nature of the CD spectra at 220 nm suggests that these structures form right-handed helices. The similarities between spectral trends also suggest that, once the transition from the 12-mer structure is complete, the higher order structures appear to adopt the same conformation. There are, however, minor differences to be noted in these structures as lack of isoellipticity at 247 nm suggests. Overall, the CD spectra of the CGG oligomers resemble a multitude of different CD spectra, none of which form a complete match to the observed CGG spectra. These spectra correspond to different structural forms that have no resemblance to each other. For example, the CD spectra of the CGG oligomers was somewhat reminiscent of a quadruplex DNA which has the CD spectral appearance of a trough-peak-trough motif.[161,162] An alternative explanation is perhaps the structure is similar to a left-handed Z-form DNA, where the alternating pyrimidine-purine-mismatch motif is sufficient to form a Z-DNA structure. The observed CD spectra also has a striking resemblance to B-form DNA with the first main peak at 280 nm severely reduced due to lack of exiton splitting within this region. As mentioned previously, because there are no reference spectra for this particular type of DNA, the CD-based structural characterization of this DNA-form remains incomplete.

The concentration-adjusted CD spectra of the complementary (CAG)n/(CTG)n duplexes, where N = A/T or C/G and n = 4, 10, 15, or 25 were consistent with B-form DNA. The (CAG)n/(CTG)n duplexes have a peak centered near 274 nm, with a slight shoulder appearing at approximately 290 nm that was not present in the 12-bp duplex (Figure 42,

Figure 42: The CD spectra of the complementary (CNG)n duplexes.

The CD spectra for duplexes of the (CNG)n/(CNG)n complementary duplexes at 25 °C in 5 mM monobasic sodium phosphate, 5 mM dibasic sodium phosphate, 0.1 mM EDTA, 100 mM sodium chloride, pH 7.0. The CD spectra are similar to that of B-form DNA structures. All CD spectra have been adjusted for duplex concentration. **A**. The CD spectra of the (CAG)n/(CTG)n duplex oligomers where n = 4 – 25 (12 – 75 base-pairs in length). **B**. The CD spectra of the (CCG)n/(CGG)n duplex oligomers where n = 4 – 25 (12 – 75 base-pairs in length).



A). The cross-over point occurred at 263 nm for the 12-mer and 30-mer and at 261 nm for all other duplexes. The troughs of the CD spectra occurred, for all duplexes studied, at 254 nm. The CD spectra suggest that the structures are forming complementary B-form DNA structures with full Watson-Crick complementarity among the matched nucleotides. The shoulders for the 45-bp and 75-bp duplexes suggest that the conformation changes as the length of the duplex increases. The nature of these changes is yet undetermined. Because of the repetitive nature of the strands under consideration, it is quite possible that the duplexes are forming slip structures, which will cause a distortion in the CD spectra signal from that of classic B-form DNA spectrum. Interestingly, the base stacking interactions, as indicated by the depth of the trough near 253 nm, and the peak near 274 nm, did not significantly increase as function of increasing length. This predicts, on the basis of base stacking interactions alone, that the duplex stability will plateau as the length increases. However, because the base stacking interactions are not the sole determinant of DNA stability, there exists the possibility that any increase in stability could be due to an increase in hydrogen bond formation.

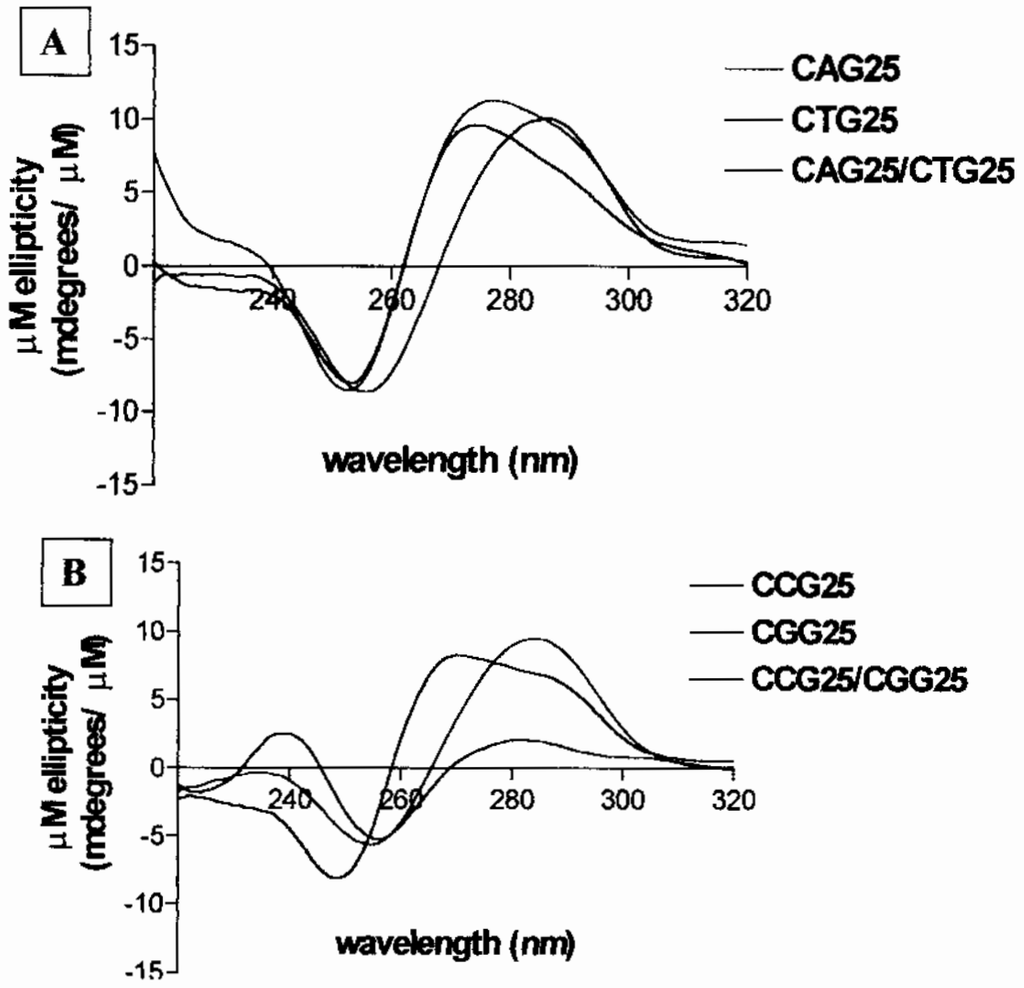
The CD spectra for the (CCG)n/ (CGG)n duplexes also suggest B-form DNA with subtle variations observed in the overall structure (Figure 42, B). For all duplexes, except the 45-bp duplex, there were two peaks centered between 274 to 270 nm, and 286 to 284 nm for the 12-bp to 75-bp duplexes, respectively. The 45-bp duplex appeared to have a single peak, centered at 280 nm, with a slight shoulder at 285 nm. The unusual CD spectra for this duplex was verified a second time, and found to be similar. There were several cross-over points that ranged from 260 nm for the 12-bp duplex, to 261 nm for

the 75-bp duplex suggesting structural variation. There was only one trough for all duplexes studied occurring from 251 nm for the 12-bp duplex to 248 nm for the 75-bp duplex. The blue-shift suggests that the nucleotide bases are becoming more exposed to the external water molecules as the length increases. This observation in combination with the blue-shift suggests that the DNA structure becomes more stable as the result of reorientation of the nucleoside bases, such that the bases become relatively more solvent accessible. As noted with the CAG/CTG duplexes, the redundant nature of these sequences raises the possibility for slip-structures, where the DNA duplex is shifted by the length of one or more repeat units. An "N+1" slip-structure, that has three solvent exposed nucleotides on each end of the duplex, could also account for the observed CD spectra. One final observation is the initial increase in the base stacking interactions, indicated by an increase in the depth of the trough near 250 nm and the increased height of the peak near 270 nm, that appears to plateau once the duplex length exceeds 45 basepairs. This result predicts that there are practical limits for the ability of these strands to accommodate additional mismatches.

An overlay plot of the CNG25 oligomers with the complementary duplexes also revealed additional trends worth noting (Figure 43). In general, the peak maximas for all solitary oligomers were red-shifted relative to the complementary duplexes, and suggests that the structures formed by the oligomers are less hydrated. As previously noted, the peak maxima of the CAG25 oligomer, which occurred near 278 nm, is the most blue-shifted of all oligomers studied and suggests that this oligomer has the highest level of hydration for

Figure 43: An overlay of the solitary oligomer and duplex DNA CD spectra.

The CD spectra for the (CNG)n triplet repeat sequences at 25 °C in 5 mM monobasic sodium phosphate, 5 mM dibasic sodium phosphate, 0.1 mM EDTA, 100 mM sodium chloride, pH 7.0. Shown below are only the overlaid spectra for the 75-mer and 75-bp DNA. All CD spectra have been adjusted for oligomer or duplex concentration. **A.** The CD spectra of (CAG)25, (CTG)25, and (CAG)25/(CTG)25. **B.** The CD spectra of (CCG)25, (CGG)25, and (CCG)25/(CGG)25.



all solitary oligomers investigated (Figure 43, A). This is consistent with a structure that contains a conformationally free mismatch that is exposed to the bulk solvent.

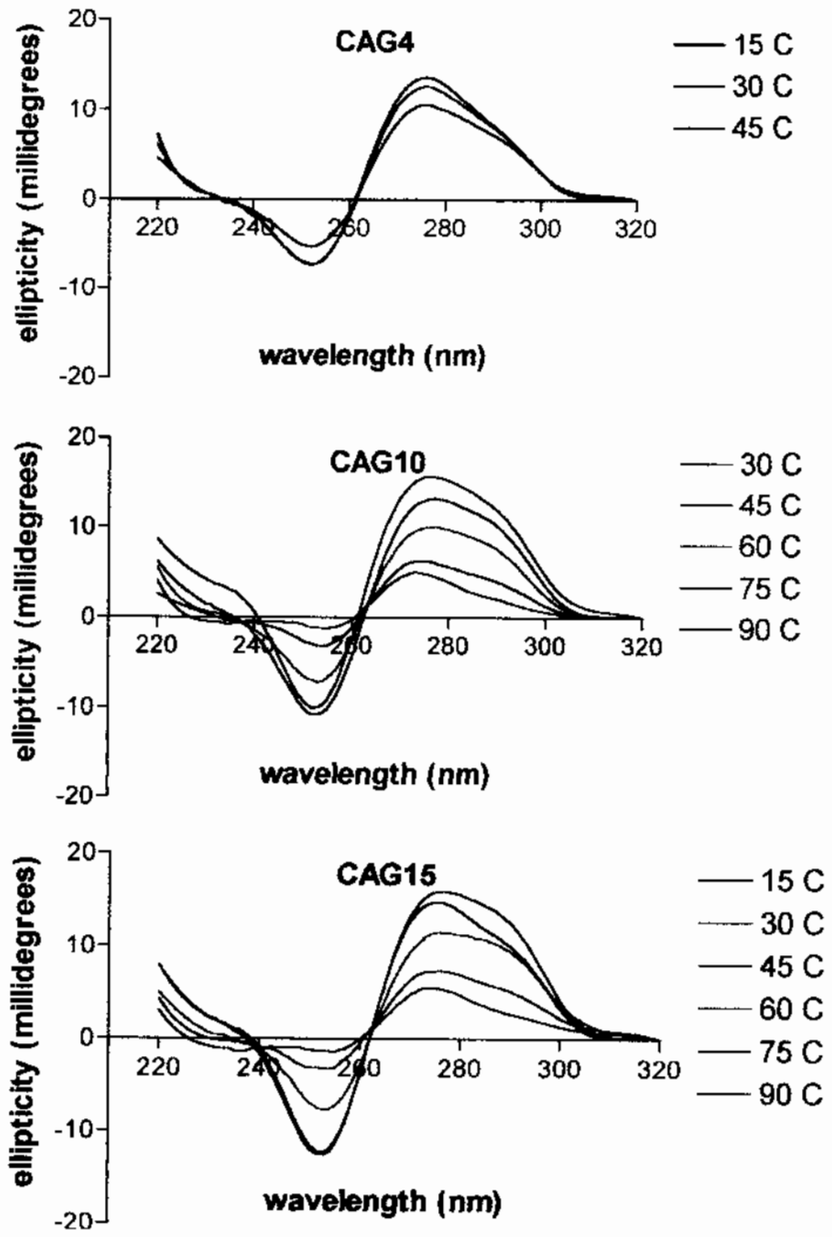
4.2.2 Thermal Melting Behavior by CD spectroscopy

The effects of thermal denaturation on the secondary structure of triplet repeat DNA oligomers (CNG)n, where N = A, C, G, or T and n = 4, 10, or 15, was studied by circular dichroism spectroscopy. For all measurements, the oligomers were in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH 7.0, and 100 mM NaCl. In general, the results show that the DNA strands hydrate and lose base stacking interactions as a function of thermal denaturation. The details of the denaturation change from as a function of oligomer sequence.

Overall, the concentration-adjusted CD spectra for the (CAG)n oligomers (Figure 44) at 15 °C were consistent with the CD spectra for previously observed samples of (CAG)n (Figure 40). This observation suggests inter-assay consistency with respect to structure. For the CAG4 oligomer there was a peak centered at 276 nm, that became blue-shifted with increasing temperature. The height of this peak decreased with increasing temperature. The crossover point occurred at 260 nm and was isoelliptic. The trough, centered at 251 nm for all temperatures, became less negative with increasing temperature. There was a shoulder, occurring near 290 nm, that became less pronounced by 45 °C. The CAG10 oligomer produced a peak centered near 275 nm that became incrementally blue-shifted to 273 nm by 90 °C. There was a shoulder occurring near the 273 nm peak, at 290 nm, that virtually disappeared at 90 °C. The height of the peak near 275 nm decreased with increasing temperature, but did not completely disappear, unlike

Figure 44: CD spectra of the thermal denaturation of (CAG)n oligomers.

The CD spectra of the (CAG)n oligomers, where n = 4, 10, or 15, were measured as a function of temperature to characterize the denaturation behavior of the oligomers.

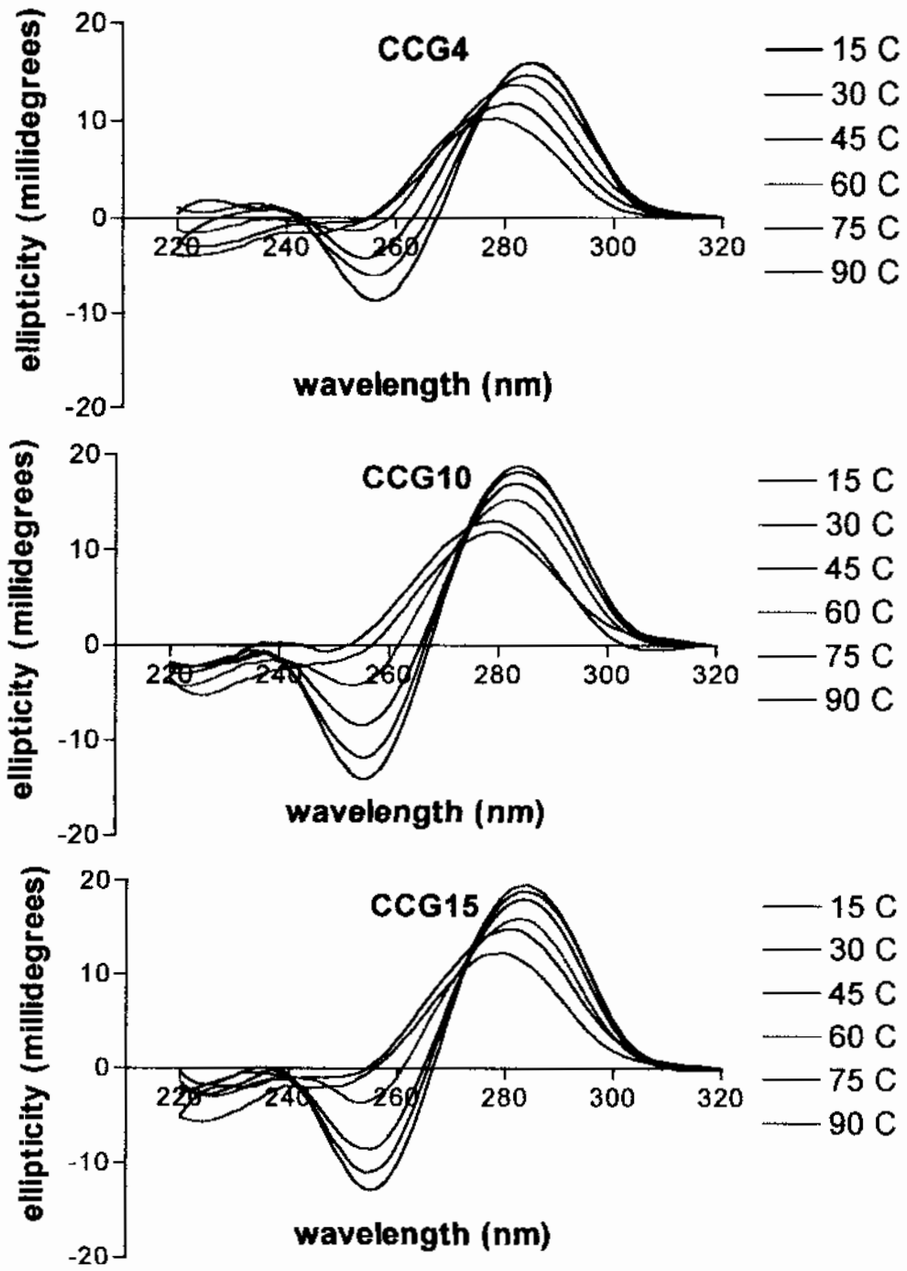


the shoulder at 290 nm. The crossover point for the spectra occurred isoelliptically, (except for the 30 °C spectrum) at 263 nm. The trough began at 252 nm at 30 °C and became blue-shifted to 253 nm at about 75 °C. The depth of the trough at 252 nm decreased until it completely disappeared by 90 °C. The CD spectra of the CAG15 oligomer produced trends that were similar to the CAG10 oligomer. There was a peak occurring at 276 nm, 15 °C that became blue-shifted to 273 nm at 90 °C. There was a shoulder occurring near 290 nm that disappeared by 90 °C. The cross-over point occurred isoelliptically at 262 nm. The trough was centered at 252 nm for all temperatures, but the depth of the trough decreased (becomes positive) beyond 60 °C. From 15 °C to 30 °C, the trough depths remained approximately constant. This data set is consistent with a DNA molecule that becomes denatured, where there is a loss of base stacking interactions and a concomitant hydration of the newly exposed hydrophobic surfaces of the nucleotides. The behavior of the shoulder at 290 nm suggests an interaction that has yet to be assigned to a structural feature. This peak could perhaps be related to the mismatched adenine nucleotide for the following reasons – the peak disappears as the temperature increases and only appears in the CAG CD spectra. The only other peak that completely disappeared as a function of temperature - the trough at 252 nm is also common to the CCG, and CTG oligomers studied.

The concentration-adjusted CD spectra for the (CCG)n oligomers were consistent with previously measured (CCG)n CD spectra, and suggests inter-assay consistency with respect to structure (Figure 45). The CCG4 oligomer had a peak, initially centered at 284 nm (15 °C) that became shifted to 277 nm at 90 °C. The cross-over occurred over the

Figure 45: CD spectra of the thermal denaturation of (CCG)n oligomers.

The CD spectra of the (CCG)n oligomers, where n = 4, 10, or 15, were measured as a function of temperature to characterize the denaturation behavior of the oligomers.



range of 267 nm at 15 °C to 254 nm at 90 °C. At 15 °C, the minima occurred near 255 nm and shifted towards 252 nm at 60 °C. By 75 °C this peak is minimal and by 90 °C the peak disappeared. The height of the peak at 284 nm decreased incrementally from 15 °C to 90 °C. As expected, the depth of the trough became positive as a function of increasing temperature. The CCG10 and CCG15 oligomers both had a peak that, at 15 °C, was centered at 284 nm, but became shifted to 277+/- 1 nm at 90 °C. Both oligomers had an isoelliptic point for all temperatures between 15 °C – 75 °C centered at 274 nm. The cross-over for the CCG10 oligomer occurred, between 253 nm and 267 nm, whereas the CCG15 oligomer it occurred between 254 nm and 266 nm. The CCG10 and CCG15 oligomers both have a trough that, at 15 °C, is centered at 255 nm but became shifted to 252+/- 1 nm at 60 °C. Between 75 °C and 90 °C, this peak disappeared. For both oligomers, the height of the peak near 284 nm incrementally decreased, and the depth of the trough near 255 nm became incrementally positive from 15 °C to 90 °C. The depth of the trough at 255 nm eventually reached zero between 75 °C and 90 °C. For all CCG oligomers studied, the data suggests that the nucleotides lose base stacking interactions (decrease in peak height, loss of trough depth), and become hydrated (blue-shifted) with increasing temperature. The CD spectra also predict, on the basis of the incremental decreases in base stacking interactions that the stability of the CCG oligomers will be low, in comparison to the CTG or CGG oligomers (Figure 46, Figure 47).

The CD spectra of the (CTG)n oligomers after correction for the oligomer concentration were consistent with previously recorded spectra and ensured that no DNA structural variation occurred as the result of sample preparation (Figure 47). For all lengths studied

Figure 46: CD spectra of the thermal denaturation of (CGG)n oligomers.

The CD spectra of the (CGG)n oligomers, where n = 4, 10, or 15, were measured as a function of temperature to characterize the denaturation behavior of the oligomers.

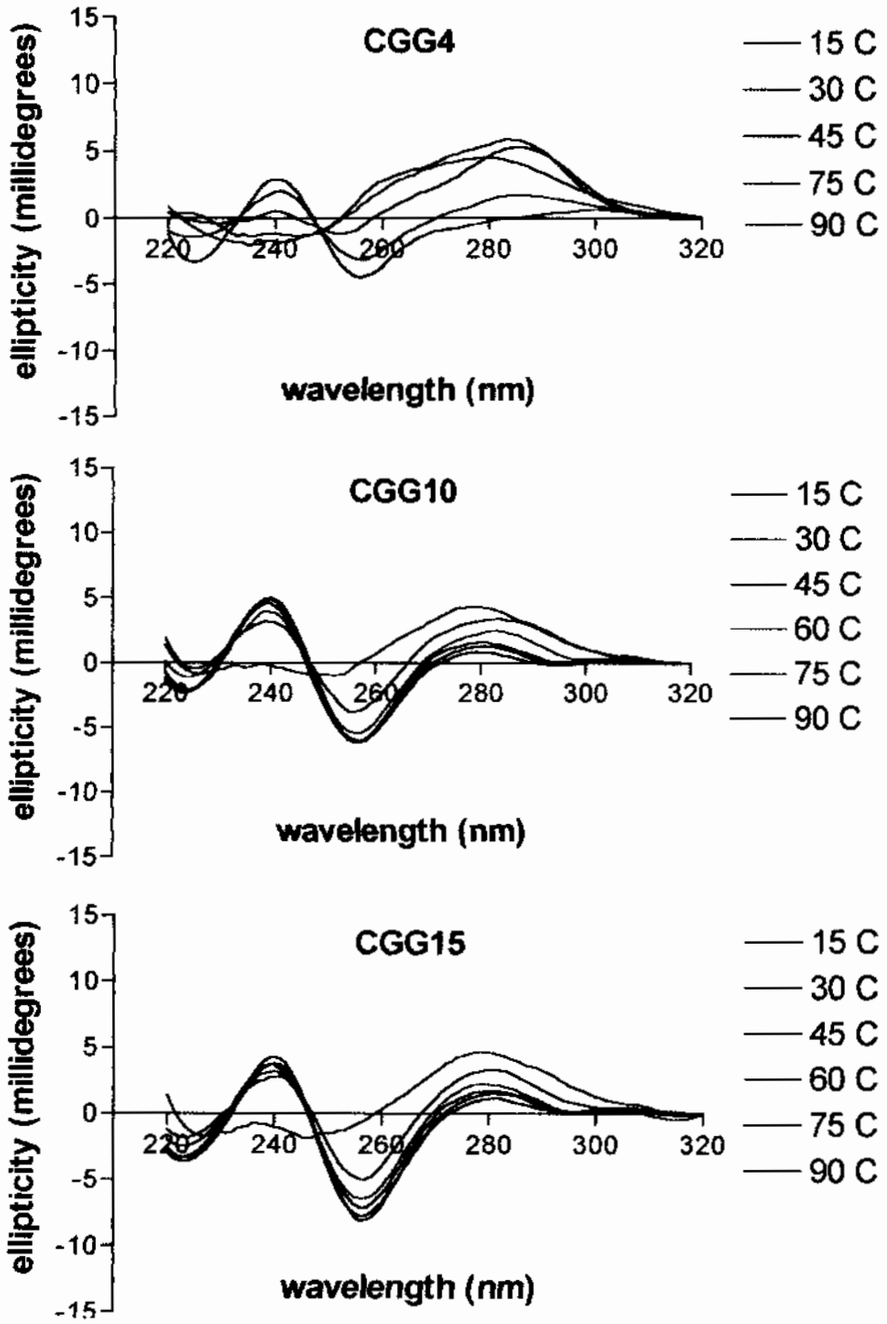
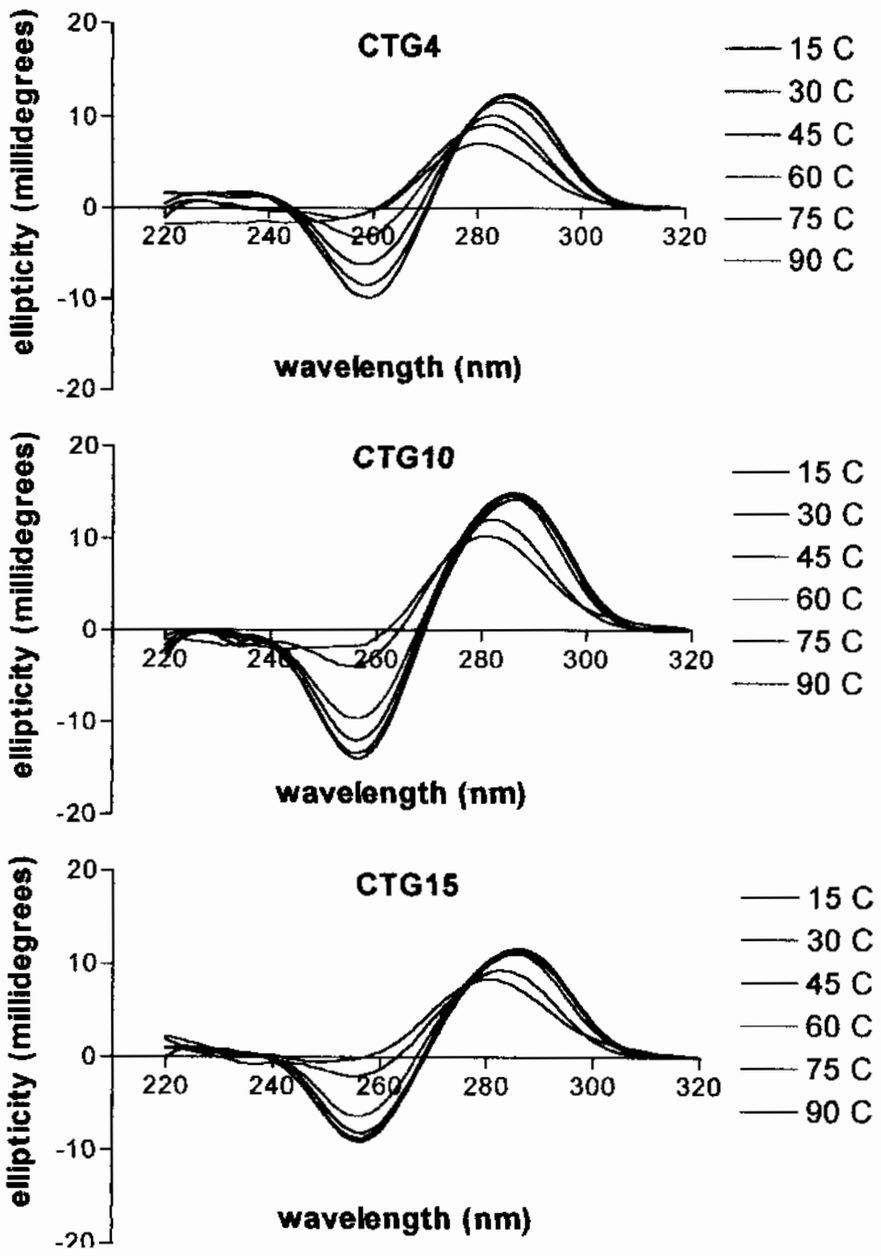


Figure 47: CD spectra of the thermal denaturation of (CTG)n oligomers.

The CD spectra of the (CTG)n oligomers, where n = 4, 10, or 15, were measured as a function of temperature to characterize the denaturation behavior of the oligomers.



at 15 °C, there was a symmetric peak centered near 286 nm that shifted to 281 nm at 90 °C. The height of the 286 nm peak remained relatively constant from 15 °C to 45 °C for CTG4, and from 15 °C to 60 °C for CTG10 and CTG15. This suggests that the DNA structure is stable within the ranges indicated, and that the oligomer does not begin to denature until relatively close to melting point. There was an isoelliptic point that occurred at 277 nm for CTG4 and CTG15. At 15 °C the symmetric minimum peak occurred at 258 nm for CTG4 and at 256 nm for CTG10 and CTG15. By 75 °C the location of the minimum shifted to 254 nm for all oligomers, and at 90 °C, the peak disappeared. For the CTG4 oligomer, the depth of this peak became incrementally positive as a function of temperature. The depth of the trough for the CTG10 oligomer remained relatively constant between 15 °C and 30 °C, but then gradually became positive until it disappeared by 90 °C. The depth of the trough for the CTG 15 oligomer produced a similar trend where the trough remained the same between 15 °C and 45 °C and then disappeared by 90 °C. The data is consistent with an oligomer having a relatively stable structure that does not begin to denature until 45 °C. When the oligomer denatures, there is a loss of base stacking interactions, as well as increased hydration of the previously buried hydrophobic surfaces of the molecule. From a structural point of view, the single isoelliptic point suggests that the CTG oligomers melt in a two-state manner.

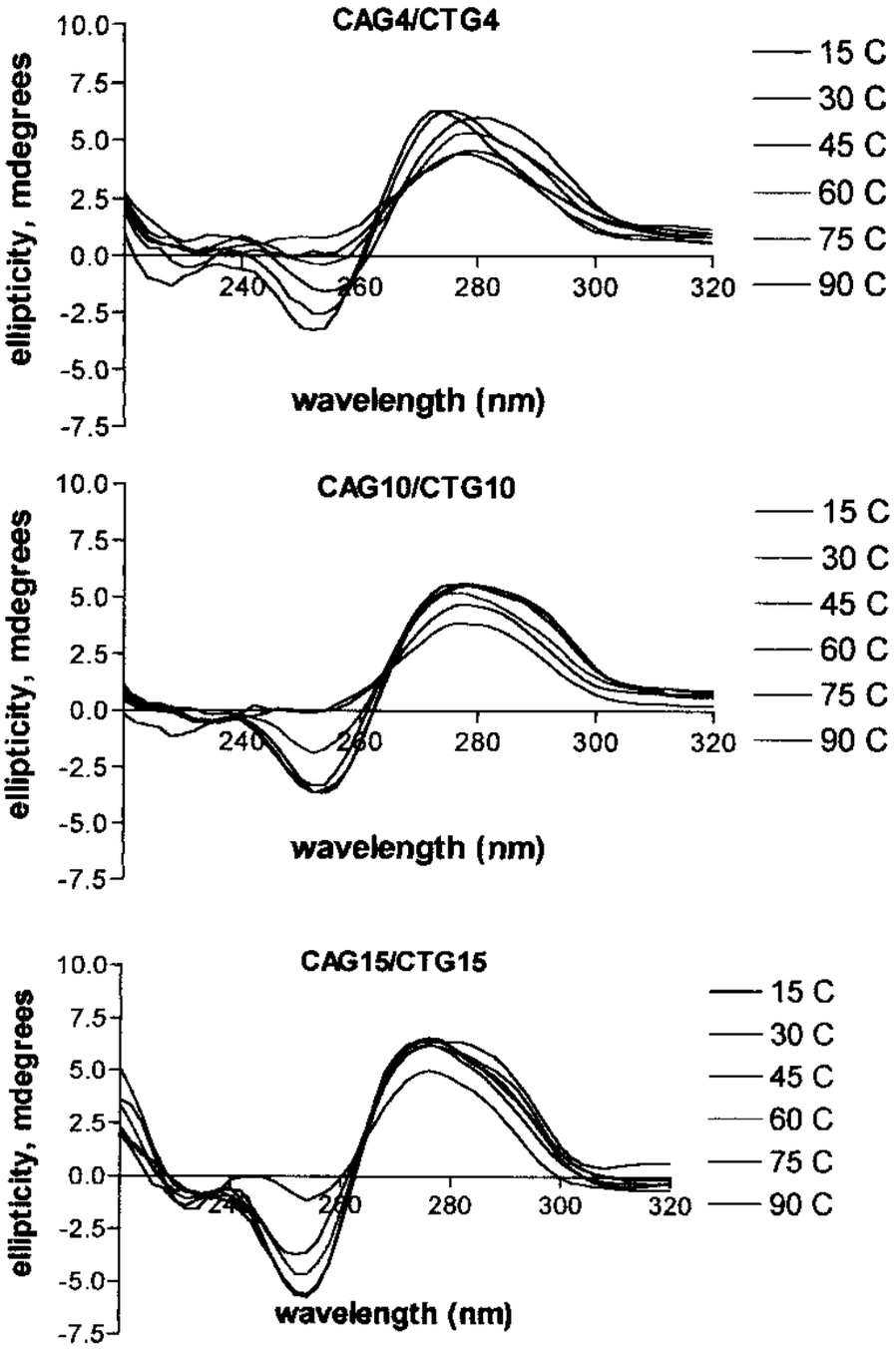
The (CGG)n oligomers displayed concentration-corrected CD spectra that were different from the CAG, CCG, or CTG oligomers studied, but consistent with previously recorded CD spectra for (CGG)n (Figure 46). For CGG10 and CGG15, there was a broad peak

centered at 280 nm (15 °C) that shifted to 279 nm at 90 °C. This particular peak displayed increased height as a function of temperature. A cursory explanation would suggest that the interaction, yet unassigned, becomes stronger with increasing temperature. However, it is possible that the trend suggests weaker interaction, because the general direction of the peak trend is towards the height of the 90 °C peak, where the strand is fully denatured - at full denaturation, the strength of the interactions should be weaker, not stronger. There is a trough beginning near 256 nm for CGG10, and 258 nm for CGG15 that shifted to 255 nm at 75 °C for CGG10, and 256 nm at 75 °C for CGG15, and by 90 °C, the trough disappeared. A second peak appeared for CGG10 and CGG15 beginning, at 15 °C, at 239 nm and subsequently shifted to 236 nm by 75 °C. This peak also completely collapsed by 90 °C. The CGG4 CD spectra produced trends that were comparable to those observed for CGG10 or CGG15. Beginning at 15 °C there was a peak, near 284 nm, that shifted to 280 nm at 90 °C. The height of this peak increased, in the range of 15 °C to 60 °C, but then decreased above 75 °C. The trough, occurred over a broad range of wavelengths, first at 255nm and then shifting towards 252 nm at 45 °C. By 75 °C, this trough disappeared. There was a second peak appearing in the CGG4 CD spectra at 240 nm for all temperatures studied. The height of this peak incrementally decreased with increasing temperature. This data suggests that, like the CTG series of oligomers, the secondary structure of the CGG10 and CGG15 oligomers remains intact between 15 °C and 60 °C, but then rapidly transitions toward the denatured form once beyond 75 °C. This predicts, that the CGG oligomers will be more stable relative to the CAG or CCG oligomers. In addition, as observed in the CD spectra of the other

oligomers, denaturation causes a loss of base stacking interactions and hydration of the previously buried hydrophobic surfaces.

The CD spectra of the (CAG)n/(CTG)n duplexes displayed similar trends to the hairpin forms of the solitary oligomers that composed the complementary strands (Figure 48). The CAG4/CTG4 complementary duplex displayed the most structural variation as a function of temperature. There was a peak initially centered at approximately 273 nm at 15 °C that became shifted to 278 nm at 90 °C. The base stacking interactions associated with this peak incrementally decreased from approximately 30 °C to 75 °C. There was an isoelliptic point only for the 45 °C through 90 °C scans that occurred at 268 nm, suggesting that the structural form of the CAG4/CTG4 duplex changed between 30 and 45 °C. There was a trough beginning at 251 nm that became shifted towards 255 nm by 60 °C. As the temperature increased, however, the location of the trough shifted backs towards the initial starting point at 252 nm. The depth of this trough incrementally increased throughout the range of 15 °C to 90 °C. In contrast to the somewhat unique spectral signal of the CAG4/CTG4 duplex, the CAG10/CTG10 and CAG15/CTG15 complementary duplexes shared many CD spectral features (Figure 48). The CD spectra contained a peak centered near 276 nm, with a slight shoulder centered between 281 nm to 285 nm. For both duplexes, the base stacking interactions remained equivalent through the range of 15 °C to 45 °C. As expected, the CAG15/CTG duplex displayed slightly greater stability, as shown by base stacking interactions that remained stable up to 75 °C. For both duplexes there was an isoelliptic point that occurred near 263 nm. The behavior of the duplexes differed slightly with respect to the trough region of the CD

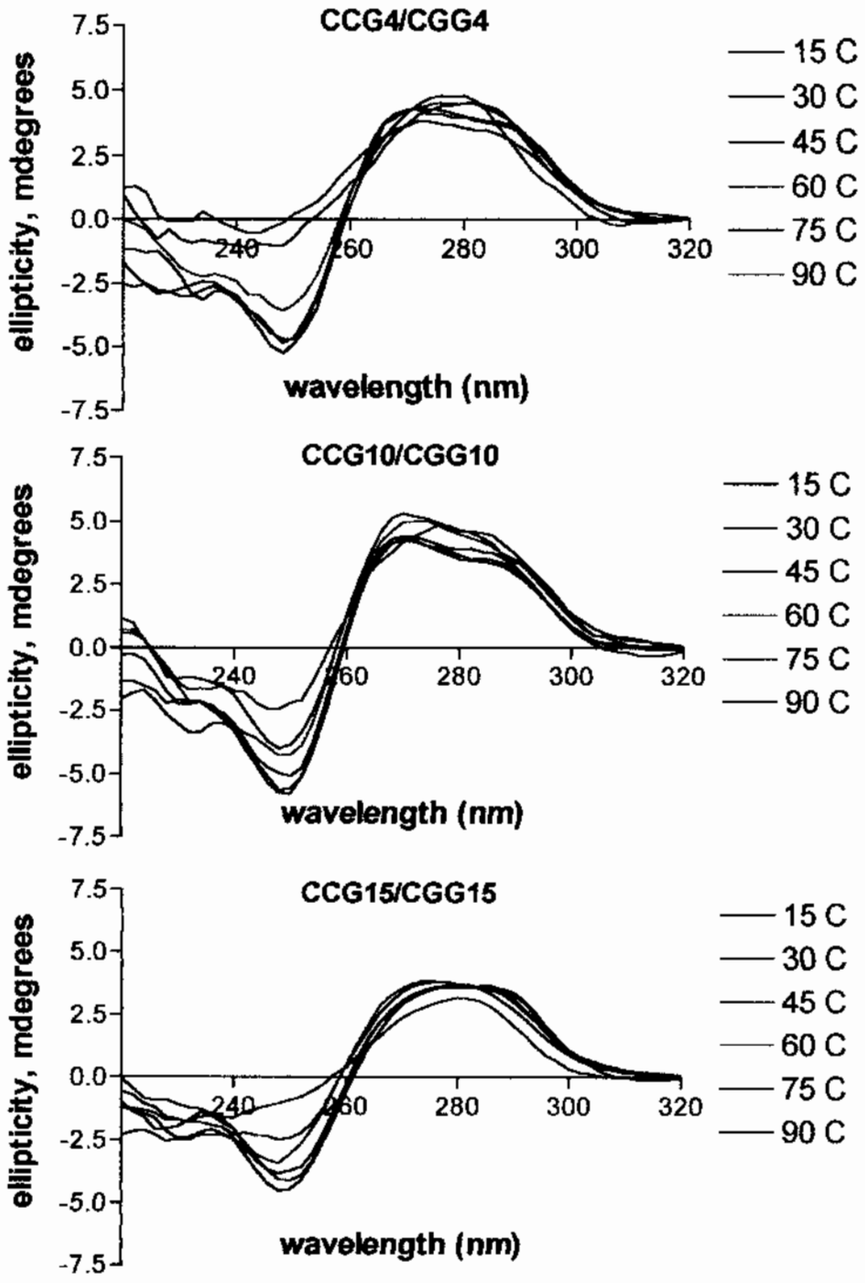
Figure 48: CD spectra of the thermal denaturation of (CAG)n/(CTG)n duplexes.



spectra. For CAG10/CTG10 the trough occurred at 253 nm and shifted slightly towards 252 nm at elevated temperatures. For the CAG15/CTG15 duplex, however, the trough occurred initially at 254 nm but then blue-shifted to 252 nm by 75 °C. Eventually the CD spectra for the 45-bp duplex shifted back to 254 nm. The base stacking interactions, as indicated by the depth of the trough at 252 nm, follow the same general trends for both duplexes. The depth of the trough decreased, indicating less interaction as the temperature increased. In general, the observed data trends are consistent with a DNA molecule that undergoes a simultaneous loss of base stacking interaction, and hydration of the nucleotide bases with increasing temperature.

The CD spectra of the complementary (CCG)n/(CGG)n duplexes, where n = 4, 10, or 15, displayed different trends across all duplexes (Figure 49). Like the CAG4/CTG4 duplex, the CCG4/CGG4 displayed the most structural variation as a function of temperature out of all the duplex lengths measured for this class. There was a peak near 272 nm with a shoulder at approximately 284 nm. For all temperatures, there was significant shifting in the location of the peak such that no clear trend as a function of temperature was apparent. The CD spectra all crossed over near 265 nm but there was no isoellipticity associated with this wavelength. The lack of isoellipticity indicates differences in structure that occurred over the temperature range studied. The base stacking interactions associated with this peak, as indicated by the height of the peak, also showed a lot of variation with temperature, such that clear trend analysis was difficult. A trough occurred at 248 nm and 15 °C that became shifted toward 242 nm at 90 °C. The depth of this trough decreased as a function of temperature beyond 45 °C, which is consistent with

Figure 49: CD spectra of the thermal denaturation of (CCG)n/(CGG)n duplexes.



the 50 °C melting transition observed for this duplex by UV-melt techniques. The peak shapes of the CAG10/CTG10 and CTG15/CTG15 duplexes were similar to each other, and that of CAG4/CTG4, however, the locations of the peak maximas and minimas differed, reflecting possibly the hydration states of the duplexes under study. For CCG10/CGG10 there was a peak occurred at 15 °C at 272 nm that became shifted towards 270 nm by 90 °C (Figure 49). Associated with this peak was a slight shoulder, observed in prior CD spectra (Figure 42) located near 287 nm at 15 °C. The location of this shoulder also shifted in the same direction as the 272 nm peaks, and finally stopped at 284 nm by 90 °C. This observation, in combination with the behavior of the 272 nm peak suggests hydration of this particular region of the nucleotide base as a function of increasing temperature. The observed height of the 272 nm peak increased as a function of temperature beyond 45 °C and showing that the base stacking interactions for this peak increased. Between the temperatures of 15 and 45 °C the observed height of this peak remained more or less constant. There was a near isoelliptic point located at 262 nm followed by a trough that began at 249 nm. The location of the trough shifted toward the blue region of the spectrum, 246 nm, at 90 °C while at the same time, the depth of this trough incrementally decreased as a function of temperature. Together the data set suggests that the CCG10/CGG10 oligomers undergo both decreased base stacking interactions, and increased hydration of the nucleotide bases as a function of temperature. As the duplexes melt the hydrophobic core of the DNA is made more accessible to the external solvent. The observed CD spectra for the CCG15/CGG15 duplex displayed trends that were consistent with DNA melting behavior (Figure 49). There was a peak at 15 °C, located at 274 nm that displayed no tendency for blue-shifting by 90 °C.

Associated with this peak was a slight shoulder, observed in prior CD spectra (Figure 42), located near 287 nm at 15 °C. The location of this shoulder, like the 274 nm peak, did not shift by 90 °C. This observation, in combination with the behavior of the 274 nm peak indicates that denaturation within this region of the nucleotide is not associated with any changes in hydration as a function of temperature. The observed height of the 274 nm peaks did not significantly change within between the range of 15 and 75 °C, but then rapidly decreased by 90 °C. This observation is consistent with retention of base stacking interactions until the relatively high melting transition is attained. There was a near-isoelliptic point located at 262 nm followed by a trough that began at 249 nm. The location of the trough shifted toward the blue region of the spectrum, 247nm, at 75°C, and completely disappeared by 90 °C. The depth of the 249 nm trough incrementally decreased as a function of temperature. Together the data set suggests that the CCG15/CGG15 oligomers undergo both decreased base stacking interactions, and increased hydration of one region of the nucleotide bases as a function of temperature. The other peak however suggests that another region undergoes only modest changes in hydration and base stacking interactions.

4.2.3 Base Stacking Interactions

In a global view, the thermal denaturation studies reveals information about the nature of the structures formed by these oligomers. The CAG oligomers had shown smaller shifts towards the blue-end of the spectrum, and have peak maxima/minima located in the blue region of the spectrum relative to the CTG, and CCG oligomers. This suggests, with the caveat that DNA sequence can change the location of the CD spectral signals, that the

CAG oligomers are more accessible to the bulk aqueous solution environment than the CCG or CTG sequences. The base stacking interactions also decrease per unit of temperature change - more for the CAG oligomers than it does for the CCG, CGG or CTG oligomers (Figure 50). This predicts, on the basis of structure, that the CAG oligomers will have lower stabilities than the other oligomers. Oligomers with intact DNA structure, will maintain the base stacking interactions. Therefore, the loss of base stacking interactions as a function of temperature can serve as a crude indicator of the melting point, and hence the stability of the DNA structure. The rank order of change, predicted by the change in base stacking interactions is CAG>CCG>CTG>CGG (Figure 50).

For a refined view of how the base stacking interactions change as a function of length, it is necessary to view the CD spectral data set adjusted not for the strand concentration, but for the nucleotide concentration. When the CD spectral data from the oligomer vs. length studies were adjusted for constant nucleotide concentration, the following trends emerged. As a function of length, the CD spectra for the (CAG)n oligomers, produced only slight changes in the height of the peak at 275 nm (Figure 51, Figure 52). The depth of the trough near 253 nm increased from the 12-mer to the 45-mer but then decreased as the oligomer length approached 75 nucleotides. These results suggests that the base stacking interactions per nucleotide initially increased with increasing oligomer length, but then decreased due to an inherent instability as function of increasing oligomer length.

Figure 50: Base-stacking interactions per strand as function of length.

The change, in terms of the average change in ellipticity at 250 nm per 15 °C is plotted as a function of oligomer length in repeat units. All CD spectra have been adjusted for oligomer concentration. The depth of the trough indicates the relative amount of base stacking interactions, and indirectly indicates the stability of the oligomer. Stable oligomers will have smaller changes to the base-stacking interactions as a function of temperature. The predicted stability order based on the slope of the line is, least stable to most stable, CAG<CCG<CTG<CGG.

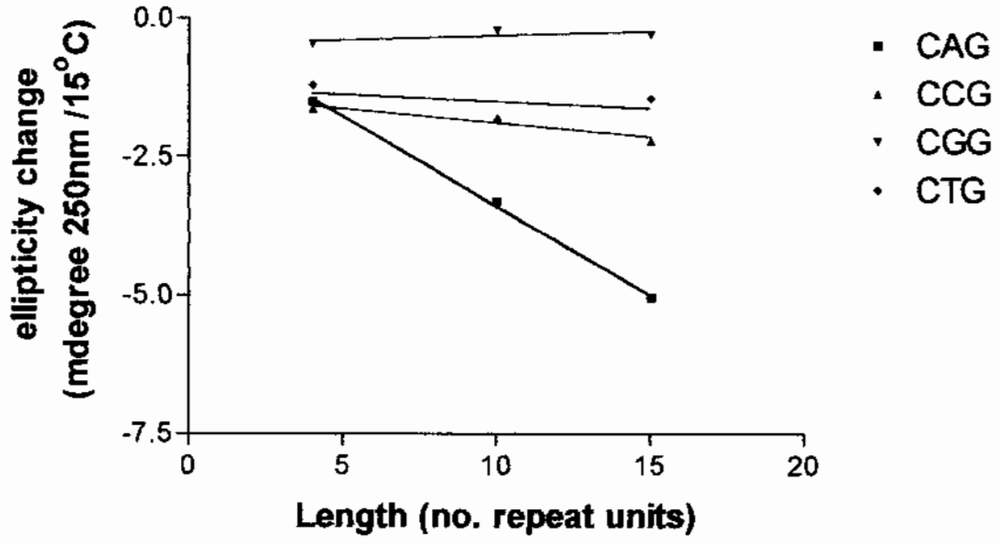


Figure 51: The CD spectra of (CAG)n and (CCG)n oligomers at constant [nucleotide].

The CD spectra for the (CAG)n and (CCG)n triplet repeat sequences at 25 °C in 5 mM monobasic sodium phosphate, 5 mM dibasic sodium phosphate, 0.1 mM EDTA, 100 mM sodium chloride, pH 7.0. The CD spectra are similar to that of B-form DNA structures. All CD spectra have been adjusted for nucleotide concentration. **A**. The CD spectra of the (CAG)n oligomers where n = 4 – 25 (12 – 75 nucleotides in length). **B**. The CD spectra of the (CCG)n oligomers where n = 4 – 25 (12 – 75 nucleotides in length).

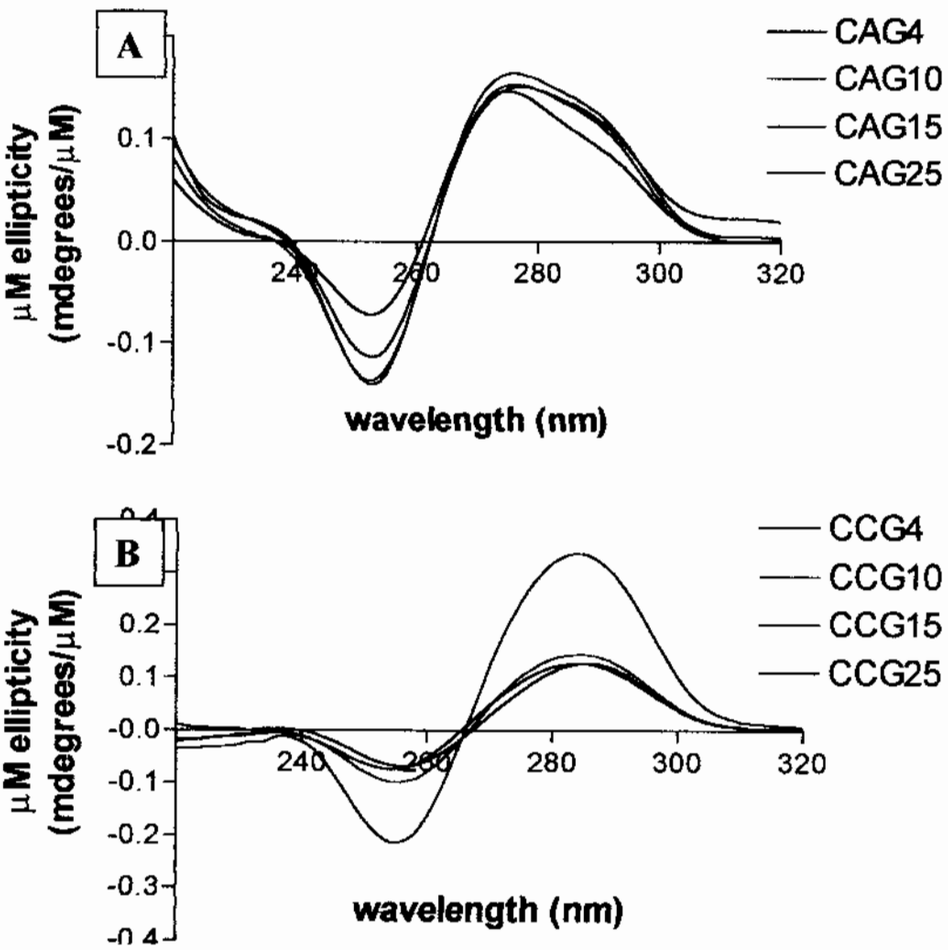
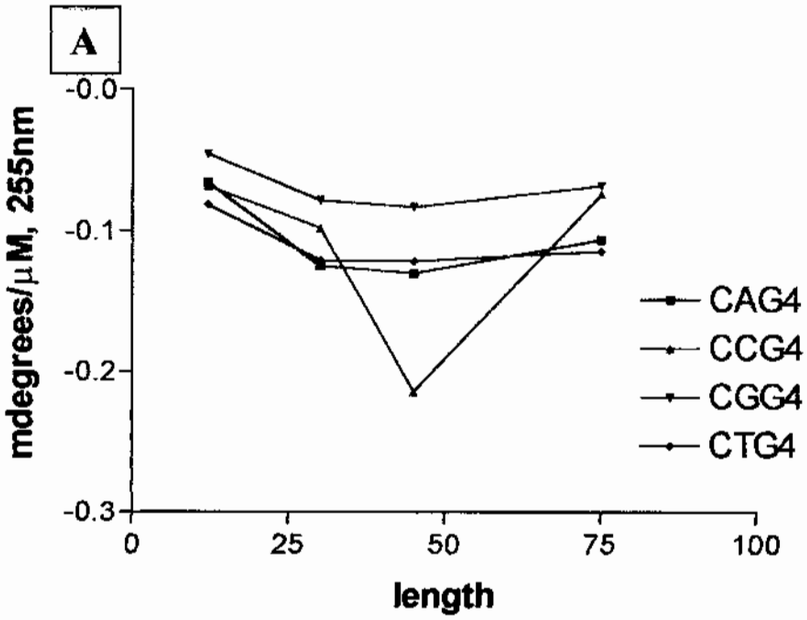
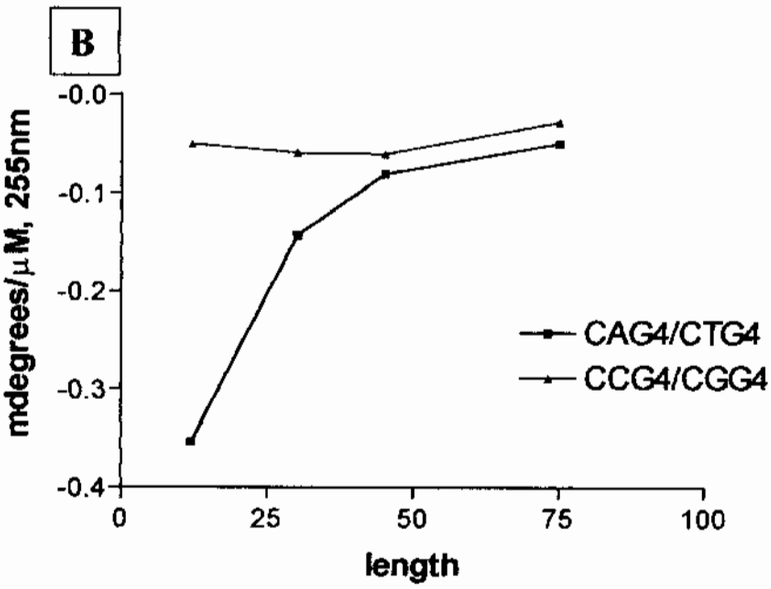


Figure 52: Base stacking interactions per nucleotide as function of length.

The ellipticity at 255 nm per is plotted as a function of oligomer length. All CD spectra have been adjusted for nucleotide concentration. The depth of the trough indicates the relative amount of base stacking interactions, and indirectly indicates the stability of the oligomer. **A**. The nucleotide concentration adjusted magnitude of the trough at 255 nm for the (CNG)n oligomers, where N = A, C, G, or T, and n = 4, 10, 15, or 25. **B**. The nucleotide concentration adjusted magnitude of the trough at 255 nm for the (CNG)n/(CNG)n duplexes, where N = A/T, or C/G and n = 4, 10, 15, or 25.





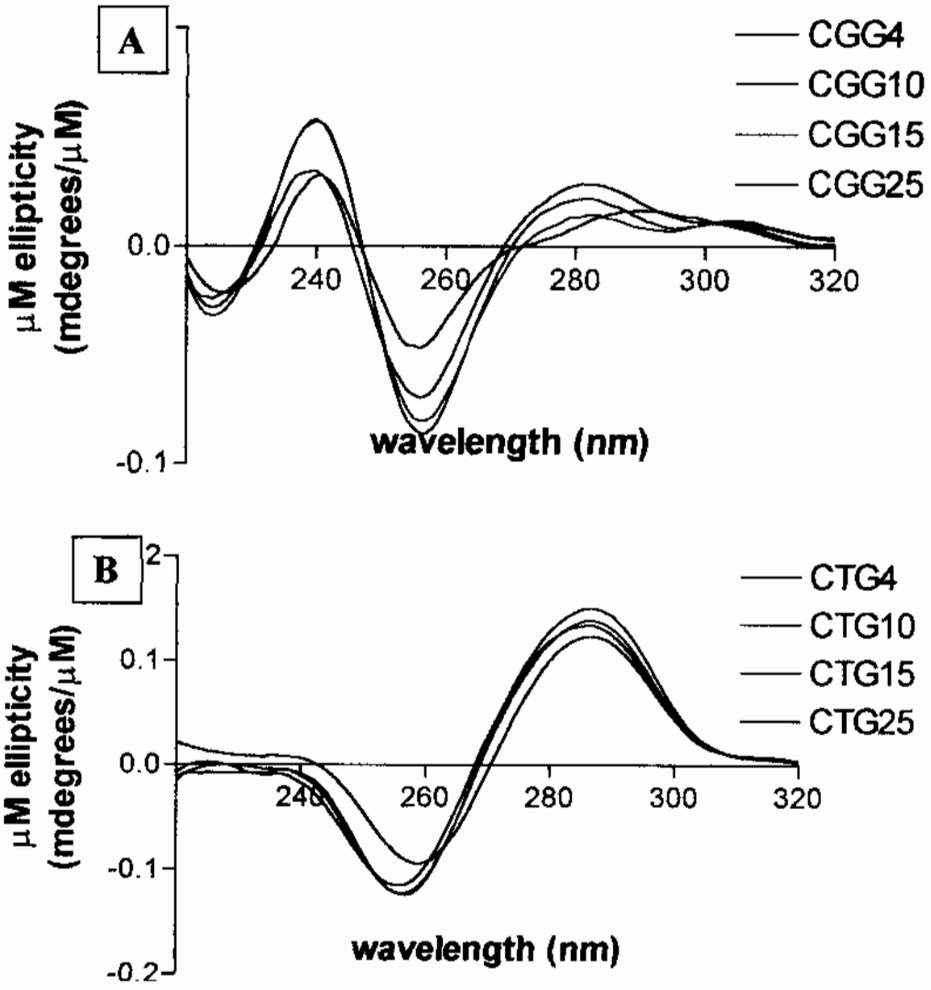
The CD spectra for the (CCG)n oligomers adjusted for nucleotide concentration displayed a similar trend (Figure 51, Figure 52). The height of peak near 284 nm increased from the 12-mer to the 45-mer, with a substantial increase occurring between the 30-mer and the 45-mer. The peak height dramatically decreased in going from the 45-mer to the 75-mer. The trough near 255 nm also displayed similar trends. The depth increased, indicating increased interactions per nucleotide, from the 12-mer to the 45-mer, but then decreased in depth in going from the 45-mer to the 75-mer.

Once adjusted for nucleotide concentration the CD spectra for the (CGG)n oligomers displayed trends that in many ways mimicked, but did not completely duplicate, those observed for the other oligomers (Figure 52, Figure 53). There was an incremental increase in the height of the peak near 280 nm that displayed no indication for decreased interactions as a function of length. The depth of the trough at 257 nm increased from the 12-mer to the 30-mer, and then incrementally decreased in going from the 45-mer to the 75-mer. A similar trend was observed for the peak occurring near 240 nm. This suggests that the base stacking interactions per nucleotide, like those observed for the other oligomers, reached a maximum at intermediate oligomer lengths.

The nucleotide-concentration adjusted CD spectra of the (CTG)n oligomers displayed subtler changes than those observed for all other oligomers, but the trends were similar (Figure 52, Figure 53). There was an increase in the peak height at 286 nm, that appeared to peak at 45 nucleotides in length. The depth of the trough also increased in a similar

Figure 53: The CD spectra of (CGG)n and (CTG)n oligomers at constant [nucleotide].

The CD spectra for the (CGG)n and (CTG)n triplet repeat sequences at 25 °C in 5 mM monobasic sodium phosphate, 5 mM dibasic sodium phosphate, 0.1 mM EDTA, 100 mM sodium chloride, pH 7.0. With the exception of the CGG series of oligomers, the CD spectra are similar to that of B-form DNA structures. All CD spectra have been adjusted for nucleotide concentration. **A.** The CD spectra of the (CGG)n oligomers where n = 4 – 25 (12 – 75 nucleotides in length). **B.** The CD spectra of the (CTG)n oligomers where n = 4 – 25 (12 – 75 nucleotides in length).



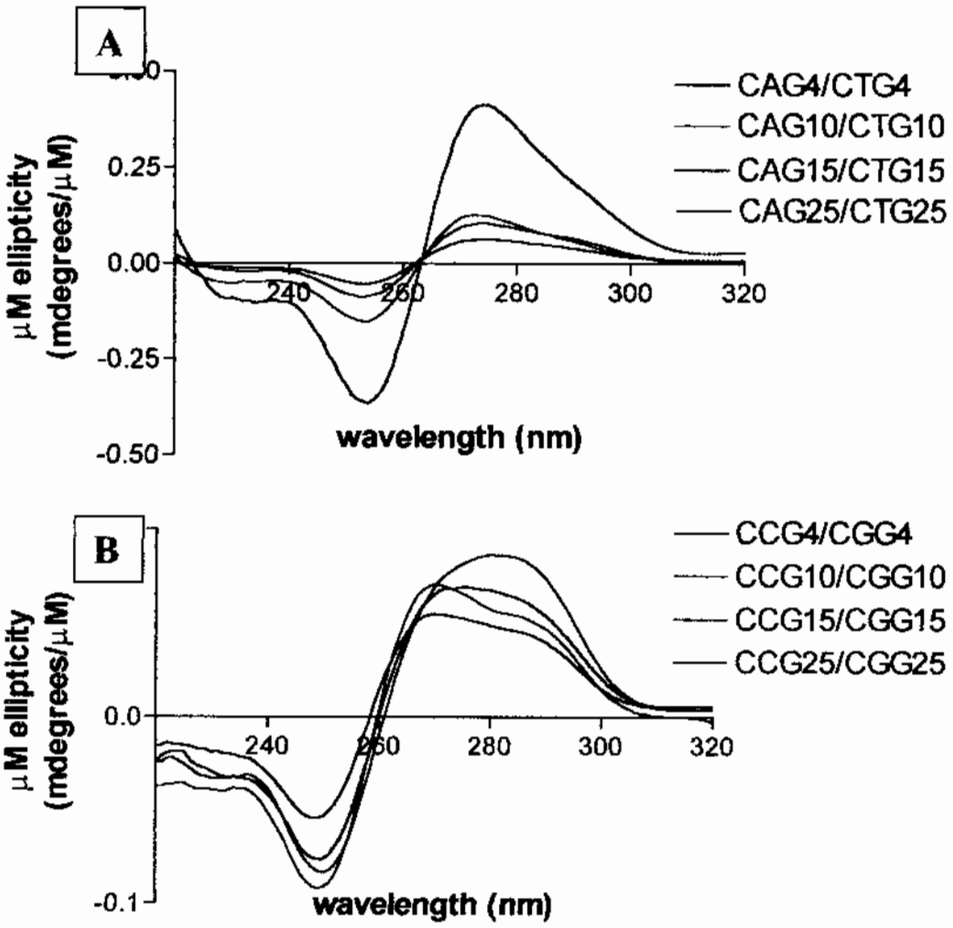
manner, and reached a peak near 45 nucleotides. The depth near 255 nm for the 75-mer was less than that of the 45-mer but more than the depth of the 12-mer.

Overall, the observed trends in the CD spectra across all oligonucleotides suggests that the base stacking interactions per nucleotide reaches a maximum somewhere around the intermediate length, 30-mer to 45-mer, within the range studied. For all oligomers studied, the 75-mer displayed more base stacking interactions than did the 12-mer, but the interactions of the 75-mers generally did not surpass that of the 30-mer or 45-mer. This suggests a lack of interaction and predicts an inherent thermodynamic instability due to the loss of base stacking interactions as the length increases. Such instability may be a necessary requirement for expansion and slippage that leads to reiterative synthesis.

The complementary duplexes displayed trends that were in part similar to that previously observed, and in others respects different than previously observed. The CD spectra of the (CAG)n/(CTG)m duplexes displayed incrementally decreased interactions as function of length (Figure 54). These interactions, shown most dramatically by the depth of the trough at 255 nm, appeared to reach a plateau as the length approached 75-bp (Figure 52). The height of the peak near 274 nm also showed a similar decrease as the length increased from 12-bp to 75-bp. These results suggest that the relative instability of CAG/CTG duplexes is linked to a loss of base stacking interactions as the length increases. The CD spectra for the (CCG)n/(CGG)n complementary duplexes displayed trends both in the peak height and trough depth that paralleled those observed for the solitary oligomers. There was some variability in the peak height that occurred between

Figure 54: The CD spectra of (CAG)n/(CTG)n and (CCG)n/(CGG)n complementary duplexes at constant [nucleotide].

The CD spectra for duplexes of the (CNG)n/(CNG)n complementary duplexes at 25 °C in 5 mM monobasic sodium phosphate, 5 mM dibasic sodium phosphate, 0.1 mM EDTA, 100 mM sodium chloride, pH 7.0. The CD spectra are similar to that of B-form DNA structures. All CD spectra have been adjusted for nucleotide concentration. **A**. The CD spectra of the (CAG)n/(CTG)n duplex oligomers where n = 4 – 25 (12 – 75 base-pairs in length). **B**. The CD spectra of the (CCG)n/(CGG)n duplex oligomers where n = 4 – 25 (12 – 75 base-pairs in length).



the 12-bp and 30-bp duplex due to a conformational change or a change in the nature of the base stacking interactions. This change rendered interpretation of the spectra between these two lengths difficult. However, the trend in the peak height from the 30-bp duplex to the 75-bp duplex showed the same type of increase. The height peaked at 45-bp then decreased as previously observed. The depth of the trough near 255 nm showed the same trend, but within this region, the trough reached the maximum depth with the 30-bp duplex. Like the solitary oligomers, the data suggests that the longer (CCG)n/(CGG)n strands, while having more base stacking interactions than the shorter strands, do not have the entire sum of possible base stacking interactions per nucleotide. Assuming the base stacking interactions increased as the sum of each additional unit of length, than the CD-spectra should remain constant when adjusted for nucleotide concentration. The CD spectra demonstrate that this is not the case, and possibly suggests that there is a loss of interactions per nucleotide as the length increases.

4.2.4 Effects of [Sodium] by CD Spectroscopy

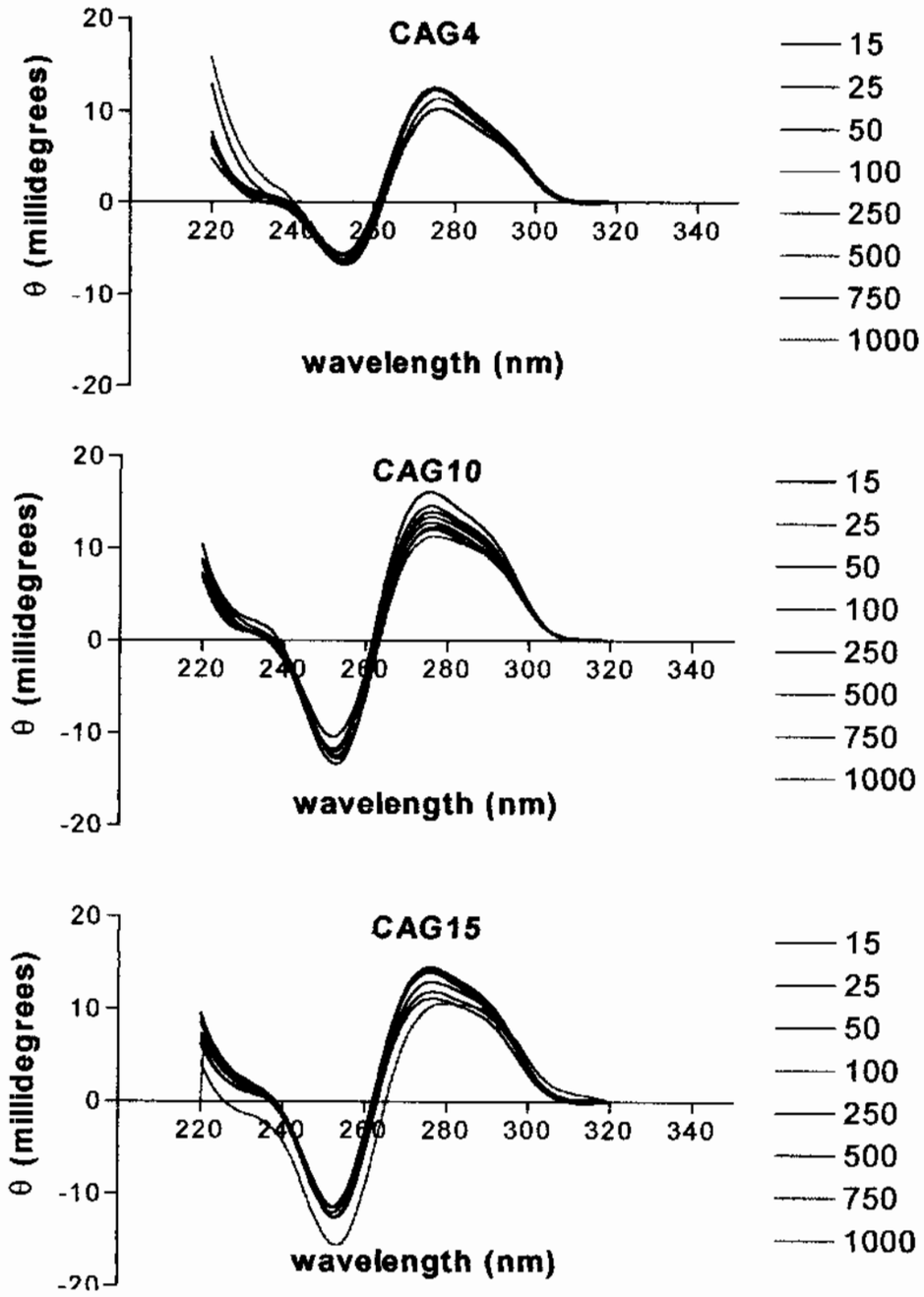
The effects of sodium concentration on the secondary structure of triplet repeat DNA oligomers (CNG)n, where N = A, C, G, or T and n = 4, 10, or 15, was studied by circular dichroism spectroscopy. For all measurements, the oligomers were in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH 7.0 with sodium chloride added to achieve a total sodium ion concentration of 15, 25, 50, 100, 250, 500, 750, or 1000 mM. For all measurements, the DNA concentrations were similar within each Na^+ titration, but different from oligomer to oligomer. The CD spectra were not corrected for oligomer concentration, because the objective in these studies was to determine the behavior of individual oligomers. In general, the effects of Na^+ concentration reveal how the

individual strands behave in response to dehydration, and indirectly provides information with respect to how these oligomers hydrate.

The (CAG)n oligomers produced CD spectra that were consistent with the spectra previously observed for the length-dependent studies at 115 mM Na^+ concentration. This observation indicates that the sample sets were consistent on an inter-assay basis and validates the sample preparation process across different procedures. The CD spectra for all CAG oligomers show one peak centered, on average, near 275 nm with a significant shoulder near 290 nm, a parallel crossover that occurred between 260 nm to 265 nm, and a trough that appeared, on average, around 252 nm (Figure 55). For all oligomers, the position of the peak became slightly red-shifted as the Na^+ concentration increased. The oligomer CAG10 displayed the smallest shift, approximately 0.5 nm that is within the experimental error of this measurement. The CAG4 oligomer displayed a modest red-shift from 275 nm at 15 mM Na^+ to 276 nm at 1000 mM Na^+. The CAG15 oligomer showed the largest shift, of 275 nm at 15 mM Na^+ to 280 nm at 1000 mM Na^+. The location of the troughs had also shown similar trends, with CAG10 having the smallest shift, from 251 nm to 252 nm, CAG4 having the next largest, from 252 nm to 253.5 nm, and CAG15 having the largest red-shift, from 252 nm to 254 nm as Na^+ is increased from 15 mM to 1000 mM. For all oligomers, the crossovers were parallel and occurred between the range of 260 nm to 261 nm for CAG4 to 262 nm to 265nm for CAG15. The crossover points followed the same general trends described for the peak and trough, namely the crossover became red-shifted as Na^+ concentration increased. Examination of

Figure 55: CD spectra of salt effects on (CAG)n oligomers.

The effects of increasing sodium ion concentration are shown for the (CAG)n series of oligomers. In all plots below, the DNA is in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH7.0, with sodium chloride added to achieve a final sodium ion concentration from 15 mM to 1000 mM (indicated in legend).



the base stacking interactions across all (CAG)n oligomers demonstrated that increasing the Na^+ concentration decreased the height of the peak near 275 nm and increased the depth of the trough near 252 nm. The CAG4 and CAG15 oligomers showed small decreases in the height of the 275 nm peak at low Na^+ concentration. Once the Na^+ concentration reached approximately 250 mM the oligomers underwent structural changes that incrementally decreased the height of the 275 nm peaks as the Na^+ concentration increased. The CAG10 oligomer however, showed an incremental decrease in the height of the 275 nm peak with increasing Na^+ concentration that began at 15 mM Na^+. Taken together, the shift towards the red end of the spectrum as a function of increasing Na^+ concentration suggests that the DNA became dehydrated as sodium ion concentration increases. This is consistent with what is known about the behavior of DNA as a function of sodium ion concentration.[163] Increasing Na^+ concentration not only dehydrates the DNA, but also allows the DNA backbone to assume a more compact structure due to shielding of the phosphate-phosphate repulsions. This more compact structure decreases solvent accessibility, to the nucleotides and at the same time, because of reorganization along the DNA backbone, shifts the alignment of the bases. Shifting of the bases from a relatively ideal, "fully hydrated DNA" type alignment to a nonideal "dehydrated DNA" alignment decreases the base stacking interactions, as observed.

The (CCG)n oligomers produced CD spectra that were consistent with previously recorded spectra at 115 mM Na^+ concentration and validated the sample preparation process across different procedures. The CD spectra for all CCG oligomers displayed one peak centered, on average, near 284 nm, a crossover that occurred between 265 nm to

267 nm, and a trough that appeared, on average, around 255 nm (Figure 56). For the CCG4 oligomer, the position of the peak and trough remained stable, at 285 nm and 255 nm respectively through the range of Na^+ concentration examined. The crossover is nearly isoelliptic, and covered the narrow range of 266 nm to 267 nm as Na^+ concentration increased. The CCG10 peak however, showed a reversal of hydration trend. At low Na^+ concentration, the peak maximum was located at 284 nm, and gradually shifted towards 283 nm at 750 mM Na^+. The spectrum then dramatically red-shifted to 285 nm at 1000 mM Na^+. The locations of the troughs remained stable, at 254 nm from 15 mM Na^+ to 750 mM Na^+, but then shifted to 257 nm at 1000 mM Na^+. In the same concentration range, the peak of the CCG15 oligomer shifted, from 284 nm to 282 nm, as the Na^+ concentration increased. The same trends were observed for the position of the trough from 255 nm at 15 mM Na^+ to 253 nm at 1000 mM Na^+. For all CCG oligomers, the height of the peak near 285 nm incrementally decreased with increasing Na^+ concentration. The behavior of the troughs however differed in a sequence-dependent manner. The depth of the CCG4 trough became deeper with increasing [Na^+]. The depth of the trough for the CCG10 oligomer remained stable up to 250 mM Na^+, then became incrementally negative with increasing Na^+ concentration. The trough of the CCG15 oligomer became incrementally positive with increasing Na^+ concentration. Overall, the data suggest that the (CCG)n oligomers, undergo dehydration with increasing salt concentration, as shown by the red-shifted spectra, and undergo decreased base stacking interactions, as shown by the decrease in the peak centered near 285nm. The loss of base stacking interactions is not the same across all (CCG)n oligomers as shown by the behavior of the troughs centered near 254 nm. This suggests

Figure 56: CD spectra of salt effects on (CCG)n oligomers.

The effects of increasing sodium ion concentration are shown for the (CCG)n series of oligomers. In all plots below, the DNA is in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH7.0 with sodium chloride added to achieve a final sodium ion concentration from 15 mM to 1000 mM (indicated in legend).

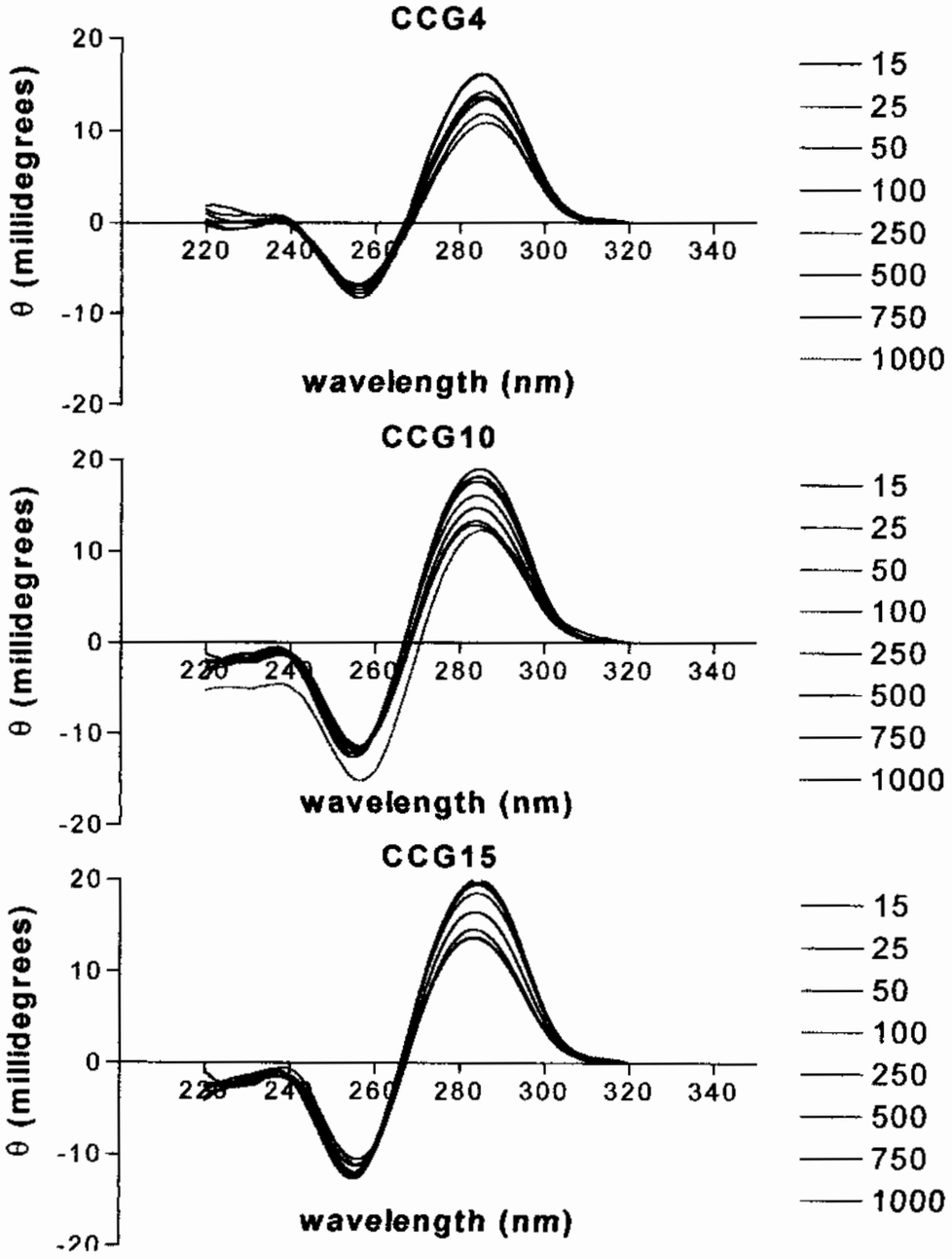
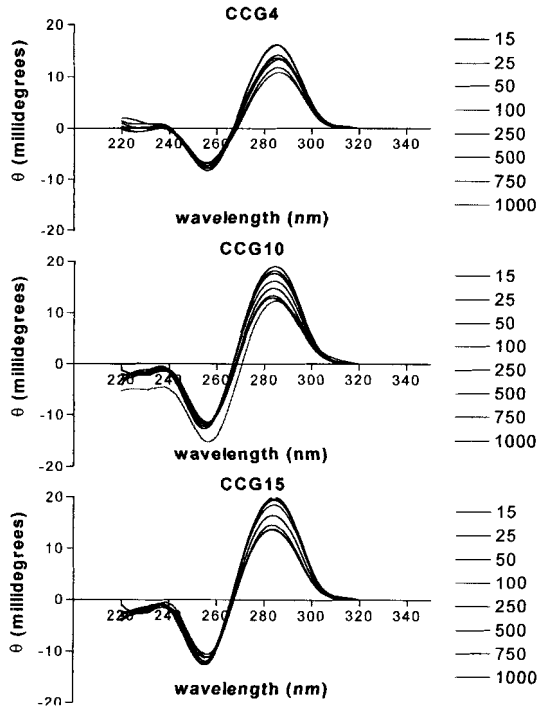


Figure 56: CD spectra of salt effects on (CCG)n oligomers.

The effects of increasing sodium ion concentration are shown for the (CCG)n series of oligomers. In all plots below, the DNA is in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH7.0 with sodium chloride added to achieve a final sodium ion concentration from 15 mM to 1000 mM (indicated in legend).

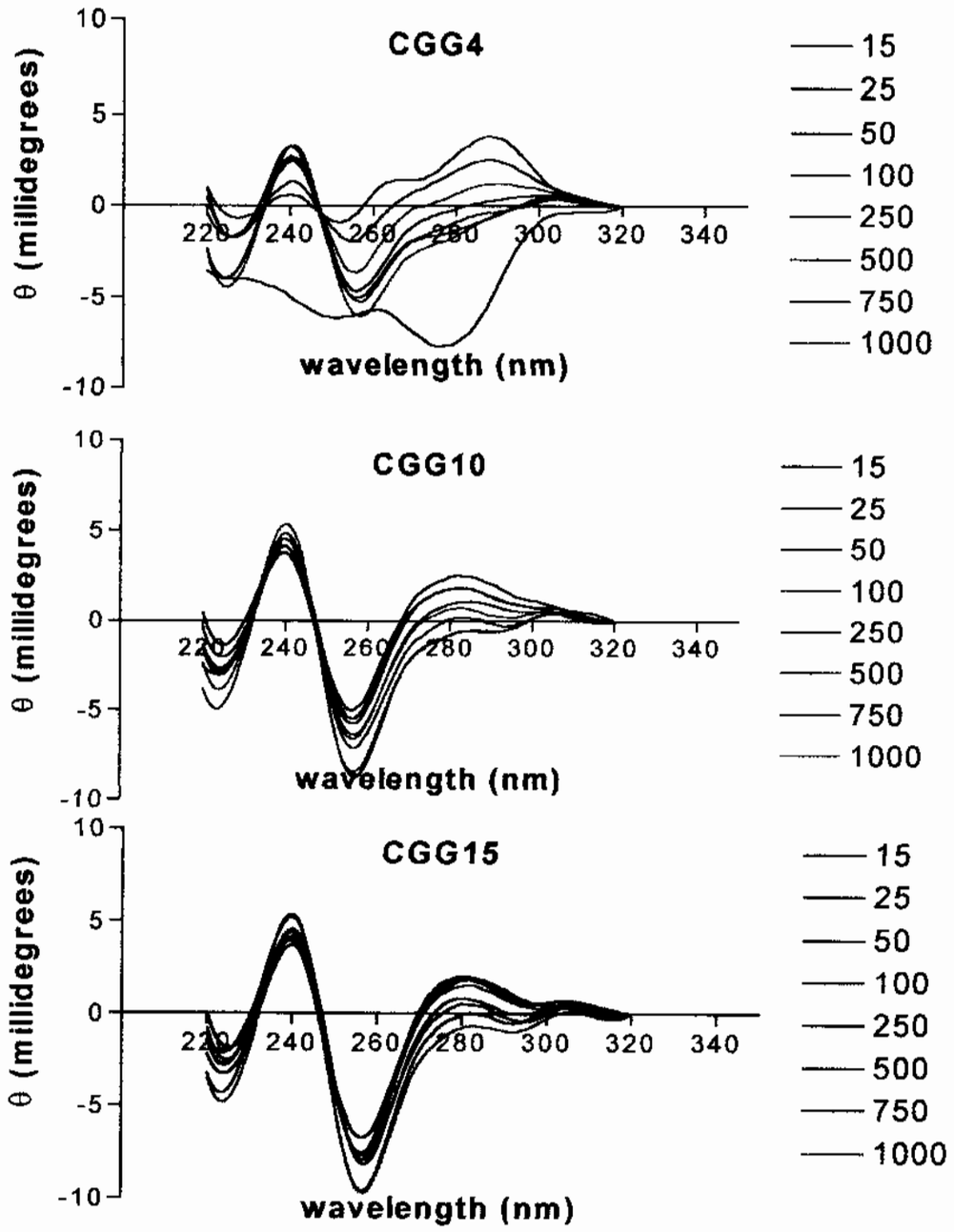


that while the DNA backbone may reorient itself to achieve a more compact structure, the orientation of the nucleotides, can vary, perhaps by *syn* versus *anti* orientation of the cytosine mismatches to achieve different degrees of stacking interactions.

The CD spectra of the (CGG)n oligomers were consistent with previously recorded spectra at 115 mM Na^+. The CD spectra for all CGG oligomers showed two peaks centered, on average, near 280 nm and 240 nm, multiple crossovers, and two troughs that appearred, on average, around 224 nm and 256 nm (Figure 57). The CGG4 oligomer undergoes a structural transition that began at 15 mM Na^+ and reached completion at 50 mM Na^+. As previously recorded the shape of the CGG4 CD spectrum in the 260nm to 280 nm region differed from that of CGG10 and CGG15 oligomers, suggesting that there is an additional structural feature associated with this molecule. The peak near 287 nm shifted from 287 nm at 15 mM Na^+ to 288 nm at 1000 mM Na^+. The height of this peak incrementally decreased as a function of increasing Na^+ concentration, and eventually became negative. The behavior of this peak is consistent with progressive dehydration of the nucleotide. There is a significant shoulder centered between 260 nm and 270 nm that began within the positive ellipticity region and became negative as the Na^+ concentration increased. The trough shifted in a manner similar to the 287 nm peak, beginning at 252 nm in 15 mM Na^+ and shifting to 257 nm by 1000 mM Na^+. The depth of the trough incrementally decreased with respect to increased Na^+ concentration. The second peak, occurring at 240 nm remained unchanged as the Na^+ concentration increased. The CGG10 oligomer is "well-behaved" and followed trends that were similar to those observed for CGG15. For CGG10, the peak at 281 nm shifted towards 283 nm as Na^+

Figure 57: CD spectra of salt effects on (CGG)n oligomers.

The effects of increasing sodium ion concentration are shown for the (CGG)n series of oligomers. In all plots below, the DNA is in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH7.0 with sodium chloride added to achieve a final sodium ion concentration from 15 mM to 1000 mM (indicated in legend).

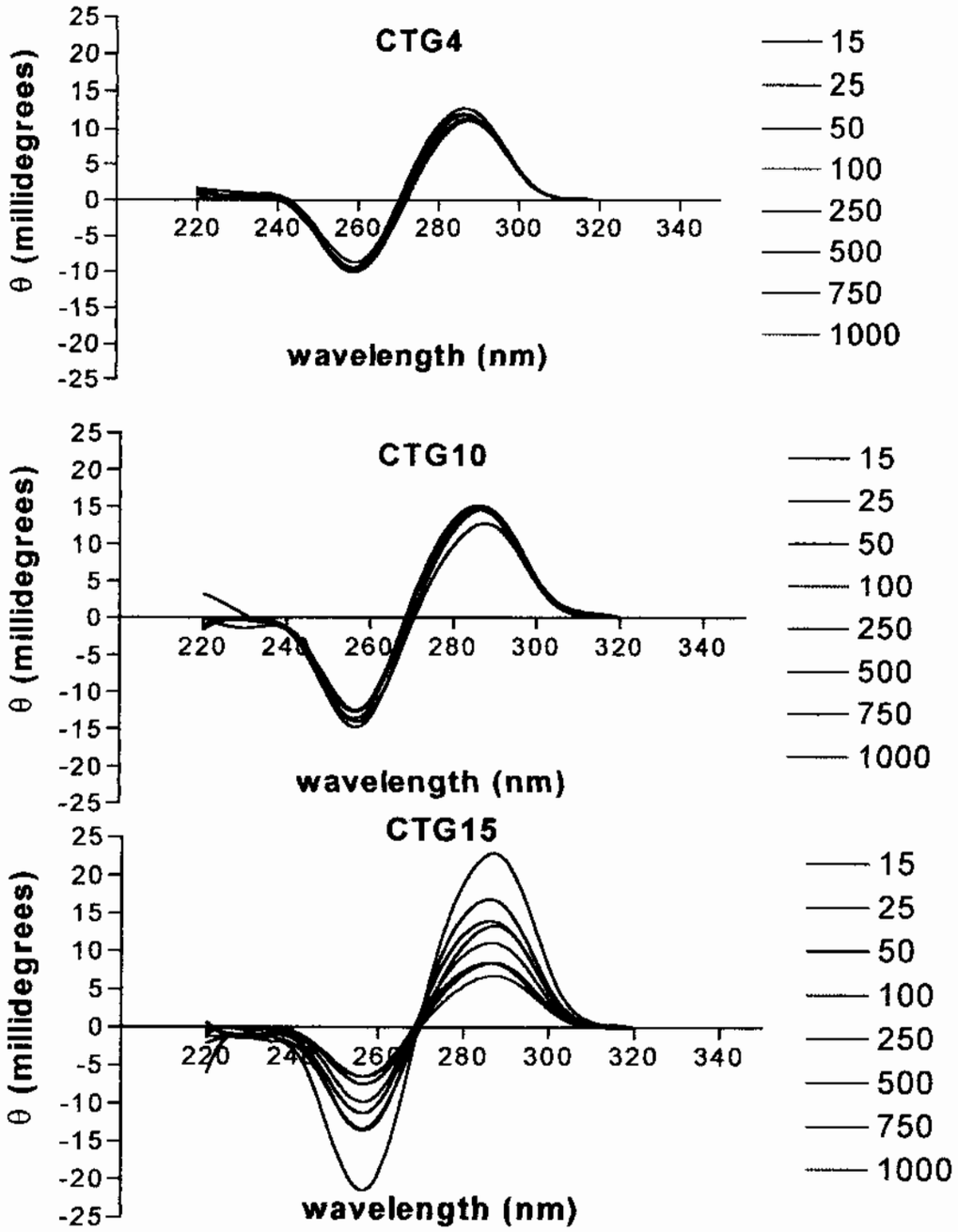


concentration increased from 15 to 1000 mM. For CGG15 a similar peak centered at 279 nm shifted towards 281 nm as the sodium concentration increased from 15 mM to 1000 mM. For both oligomers, the height of this peak incrementally decreased as function of increased sodium concentration, and eventually became negative at 750 mM Na^+. The trough occurring at 255 nm for CGG10 in 15 mM Na^+ shifted to 257 nm in 1000 mM Na^+. Likewise, the trough for oligomer CGG15, occurred initially at 256 nm in 15 mM Na^+, shifted to 251 nm in 1000 mM Na^+. For both oligomers, the depth of the trough became more negative with increased Na^+ concentration. Overall, the data suggests that the CGG oligomers follow the same trends observed for the other (CNG)n oligomers. Namely, the DNA becomes dehydrated as sodium ion concentration increases. This dehydration produces a more compact DNA structure, as suggested by the red-shift of most of the major peaks observed. Compaction of the DNA backbone due to minimization of the phosphate-phosphate repulsion causes a realignment of the nucleotide bases. This reorients the bases such that some of the bases lose base stacking interactions, as shown by the decrease in the 281 nm peak, whereas other portions of the molecule gain interactions, as shown by the gain in the 240 nm peak.The (CTG)n oligomers produced CD spectra that were consistent with those previously observed at 115 mM Na^+ and validated the sample preparation process across different procedures.

The CD spectra for all CTG oligomers showed one peak centered, on average, near 285 nm, a crossover that occurred between 266 nm to 271 nm, and a trough that appeared, on average, around 257 nm (Figure 58). Overall, the most impressive feature for all CTG CD spectra is the relative lack of change for the CTG4 and CTG10 oligomers. This

Figure 58: CD spectra of salt effects on (CTG)n oligomers.

The effects of increasing sodium ion concentration are shown for the (CTG)n series of oligomers. In all plots below, the DNA is in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH7.0, with sodium chloride added to achieve a final sodium ion concentration from 15 mM to 1000 mM (indicated in legend).



observation is contrasted against the major changes observed for the CTG15 oligomer. The location of the peak for the CTG4 oligomer shifted slightly from 285 nm at 15 mM Na^+ to 287 nm at 1000 mM Na^+. The crossover point also displayed a narrow range, from 270 nm at 15 mM Na^+ to 271 nm at 1000 mM Na^+. The location of the trough demonstrated no Na^+ concentration dependence, and remained at 260 nm for all concentrations studied. The base stacking interactions were small and can be considered as nearly equivalent across the entire range of Na^+ concentrations. The CD spectra of the CTG10 oligomers contained a peak that gradually shifted from 285 nm at 15 mM Na^+ to 287 nm at 1000 mM Na^+. The crossover occurred through a narrow range, from 268 nm to 270 nm with increasing Na^+ concentration. The location of the trough was independent of Na^+, and occurred at 265 nm for all concentrations. The height of the peak near 285 nm remained unchanged in the range of 15 mM Na^+ to 500 mM Na^+, but then decreased at 1000 mM Na^+. The depths of the troughs gradually became more positive beyond 250 mM Na^+. The CD spectra for the CTG15 oligomers, while similar in shape to previously recorded spectra, shown trends that did not duplicate the trends observed for oligomers within this series. The location of the peak maxima occurred between two wavelengths as a function of concentration. The maxima were located at 286 nm between 15 mM and 250 mM Na^+, but then shifted to 287 nm between 500 mM and 1000 mM Na^+. There was no gradual shift in the location of the maxima. The crossover point was isoelliptic, and occurred at 266 for all Na^+ concentrations. The location of the trough, at 256 nm, displayed no concentration dependence. The base stacking interactions, shown by the height of the peak and the depth of the trough showed unsual trends that were not duplicated with CTG10 or CTG4. The height of the peak

incrementally decreased from 25 mM to 250 mM Na^+, but then increased at 500 mM Na^+. At 500 mM Na^+ the height of the peak incrementally decreased, but the observed decreases did not fall below the lowest level achieved by the 250 mM sample. A parallel trend was also observed for the behavior of the trough. The depth became more positive between 25 an 250 mM Na^+, but then sharply became more negative at 500 mM Na^+. At 500 mM Na^+ the depth began to assume incremental and positive increases, but the total movement of the increase did not surpass the 250 mM Na^+ sample. Overall, the data suggest that the CTG4 and CTG10 oligomers undergo subtle structural changes whereas the CTG undergoes major structural changes as a function of Na^+ concentration. The data suggest that the CTG4 and CTG10 oligomers are stable to environmental solution changes caused by increasing Na^+ concentration. Apparently the dehydration and or compaction brought about by increasing sodium ion concentration does not appreciably affect the sum total of the base stacking interactions and, by inference, the orientation of the bases. A different trend is observed for the CTG15 oligomer, which undergoes marked changes in structure as Na^+ is titrated from 15 mM to 1000 mM. The changes may be related to the thymine mismatches that may possibly undergo a *syn/anti* reorientation in response to Na^+ concentration. Once the reorientation is complete, the DNA backbone assumes a more compact structure and, as the Na^+ concentration increases, the optimal alignment of the bases is distorted resulting in decreased interactions.

4.2.5 Concentration Dependence of the T_m

The CD spectral information indicated that the CAG, CCG and CTG oligomers formed DNA structures that were similar to, if not consistent with, B-form DNA. However, the structural type – hairpin or mismatched duplex could not be assigned due to the self-complementary nature of these sequences under consideration. Both structures could be easily formed and neither form could be readily ruled out on the basis of CD spectra alone, because no such reference spectra exist. Complicating matters further, the CGG oligomers produced unusual CD spectra that could not be readily assigned to a particular structural class of DNA. The CD spectra of the CGG oligomers were similar to that of quadruplex DNA, while at the same time had elements of a variety of different DNA-structural forms. To clarify the type of structure formed, the concentration dependence of the melting transition, T_m, was measured. Measurement of the concentration dependence could determine the molecularity of the DNA dissociation, and, by deduction, eliminate one or more of the possible structural forms under consideration. If, for example, the molecularity of association, n, was equal to 2, then the hairpin (n = 1) and the quadruplex (n = 4) structures could immediately be eliminated from consideration.

The concentration dependence of the melting temperature was measured for all oligomers over a 20 – 100 fold range in DNA concentration, in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH7.0, 100 mM NaCl. The DNA concentration was from 1.6 to 237 μM for CAG4, 2 to 54μM for CAG10, and 1 to 3 μM for CAG15. The DNA concentration was from 1 to 365 μM for CCG4, 3 to 129 μM for CCG10, and 2 to 69 μM for CCG15. The DNA concentration was from 1 to 146 μM for CGG4, 0.79 to 23 μM for CGG10,

and 2 to 117 μM for CGG15. Finally, the concentration range was from 1 to 43 μM for CTG4, 1 to 8 μM for CTG10, and 0.6 to 4.6μM for CAG15. The absorbance as a function of temperature was measured at the optimum UV-difference wavelength for each oligomer under study. The average wavelength, to give an indication of where this wavelength was located in the UV spectrum, was 255 nm. The melting temperature was determined by converting the absorbance versus temperature plot into an α-plot as previously described. The melting transition temperature was plotted as a function of the natural log of the DNA concentration. The slope of the linear regression line that was fit to the data set determined "n", the molecularity of association (Table 17). Melting transitions with a molecularity of two or more will be affected by the DNA concentration due to shifts in the equilibrium constant in a concentration dependent manner. Associations of one, such as the melting of a self-associating hairpin, will have melting transitions independent of the DNA concentration, because the equilibrium constant does not factor into the equation (Equation 23).

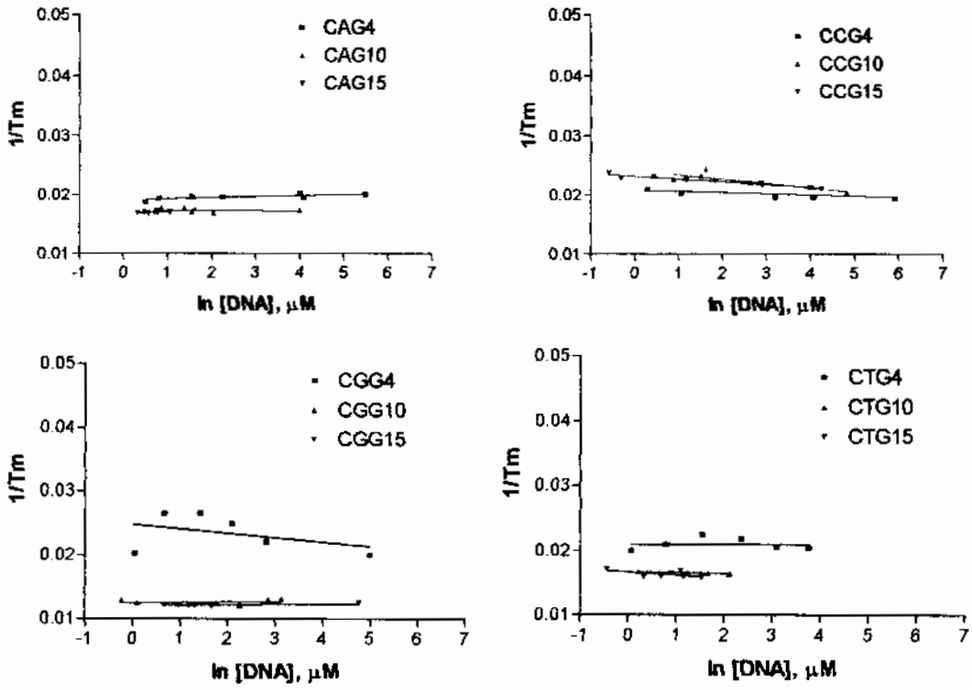
The (CAG)n series of oligomers, where n = 4, 10, or 15 showed near zero slopes indicating that the molecularity of the transition is 1 (Figure 59, Table 17). The (CCG)4 oligomers showed similar trends. The slopes of the lines were near zero, suggesting a molecularity of one. The slopes of the linear regression lines for the CCG10 and CCG15 oligomers indicated a slight downward trend (Table 17). While the slope is not significant enough to suggest that this oligomer is associating with higher order molecularity, it does suggest the possibility for other interactions occurring with this oligomer. The molecularity is therefore "conditionally" set at one. The slopes of the

Table 17: Linear regression fit parameters for the [DNA] dependence of the T_m.

Oligomer	m	b	r2
CAG4	-0.0002	0.019	0.53
CAG10	0.0004	0.0174	0.22
CAG15	0.0003	0.0168	0.2
Oligomer	**m**	**b**	**r2**
CCG4	0.0002	0.021	0.28
CCG10	-0.0007	0.024	0.64
CCG15	-0.0005	0.0233	0.92
Oligomer	**m**	**b**	**r2**
CGG4	0.0007	0.0248	0.1
CGG10	7E-05	0.0125	0.17
CGG15	-9E-05	0.012	0.63
Oligomer	**m**	**b**	**r2**
CTG4	2E-05	0.0209	0.14
CTG10	-0.0002	0.0167	0.8
CTG15	0.0005	0.0165	0.32

Figure 59: DNA concentration dependence of the melting transition.

The concentration dependence plots for all oligomers, except CGG4 and CTG4, display near-zero slopes ,that suggest unimolecular transitions.



CGG10 and CGG15 plots were near zero also suggesting that the association molecularity is one (Table 17). The plot for CGG4 however, displays a high degree of variability that can be interpreted in one of two ways. First, that the trend for this line is biphasic, where one half contains a positive slope and the other half contains a negative slope. Such a line predicts that the CGG4 oligomer forms two different structures in a concentration dependent manner. Alternatively, the trend of the melting temperatures could be biased towards large errors due to the small difference between the denatured and native extinction coefficients for this oligomer. Errors made to the pre and post transition baselines for plots with shallow slope can have enormous effects on the measured T_m. Both interpretations were equally valid and required additional experiments to rule out the possibility of quadruplex or other structures. The plots for the CTG10 and CTG15 oligomers showed near-zero slopes indicating that the melting temperature is independent of the concentration, and the association has a molecularity of one (Table 17). The CTG4 oligomer however, displayed a trend similar to that observed for CGG4. Either the plot was biphasic, or the error within the measurement is large due to the manner in which the melting transitions were determined.

In summary, the oligomers CAG4, 10 and 15; CCG4, 10, and 15; CGG10 and 15; CTG 10 and 15 all showed DNA concentration dependence lines with slopes that are near-zero, suggesting that these oligomers undergo unimolecular transitions. The trends for oligomers CGG4 and CTG4 were not as clear, warranting further investigation of the structures formed by these oligomers.

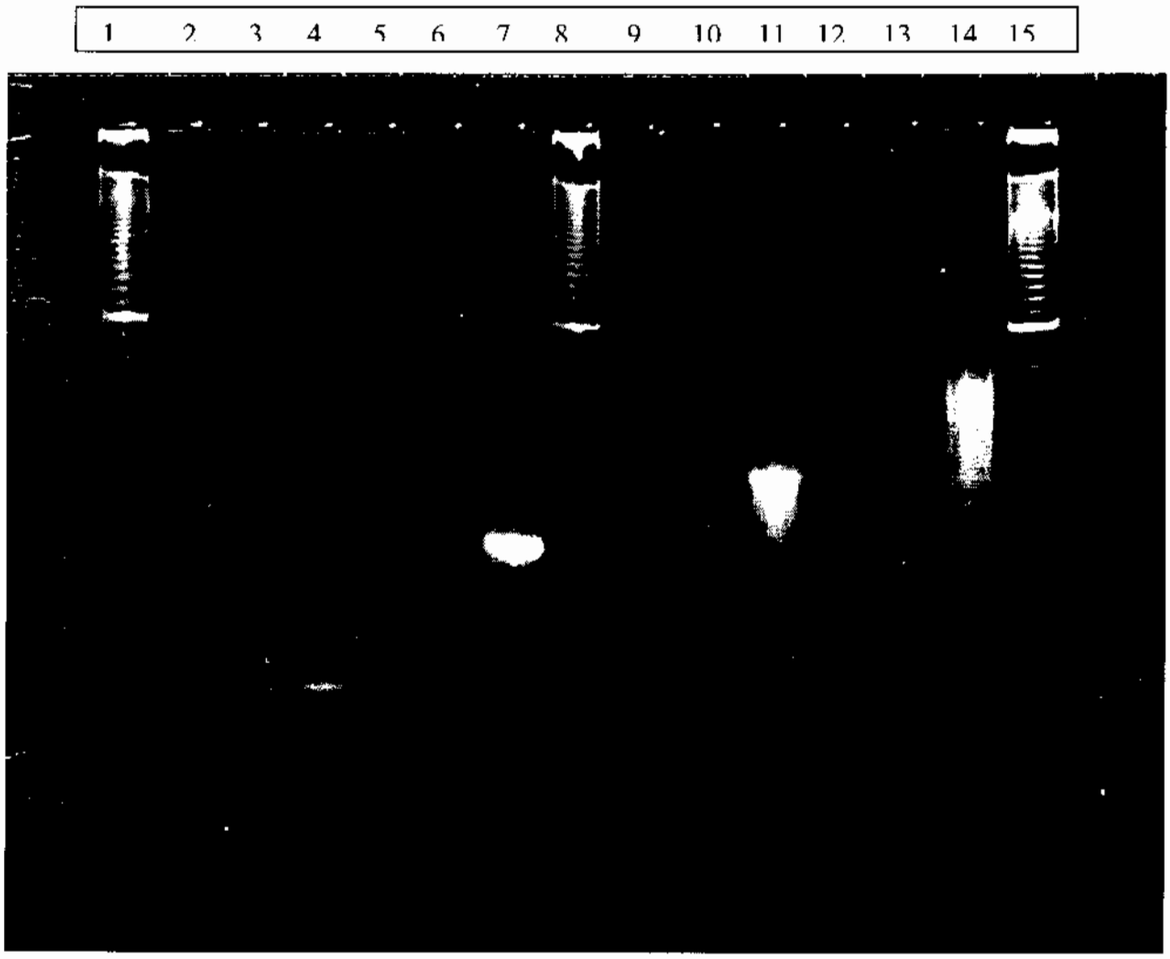
4.2.6 Native Gel Electrophoresis

The CNG triplet repeat oligomers, with lengths equal to 12, 30, 45, or 25 nucleotides long were examined by native polyacrylamide gradient-gel electrophoresis to determine the size and molecularity of the structures formed. All oligomers, at concentrations spanning the range of 100 to 200 μg/ml were previously equilibrated in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH 7.0, 100 mM NaCl for a minimum of 48 hours prior to measurement as previously described. The concentration of DNA that was most likely to produce secondary structures, if any could form, was selected with the constraint that the amount of DNA applied to the gel had to be within the linear range of this technique. Because both duplex and quadruplex structures are favored at high DNA concentrations, the range of DNA concentrations tested was between 4.2 - 8.4 μM for 75-mers, and 29 - 58 μM for 12-mers. In addtion, each oligomer was analyzed in parallel with the complementary duplex for that sequence.

Analysis of the (CAG)n and (CTG)n oligomers in conjunction with the corresponding (CAG)n/(CTG)n duplexes displayed single bands migrating with an apparently faster mobility than the corresponding duplex (Figure 60). The bands for the CAG4 and CTG4 oligomers migrated with an approximate size of 10 and 10.5 bp respectively, whereas the complementary duplex migrated with an apparent size of 12-15 bp. The CAG10 and CTG10 oligomers migrated with a position on the gel that corresponded to approximately 24 and 23 base pairs, respectively, whereas the band for the complementary duplex appeared at approximately 29 base pairs. The CAG15 and CTG15 oligomers comigrated at approximately 38 base pairs in size, while the complementary duplex migrated at a

Figure 60: Native gel electrophoresis of (CAG)n, (CTG)n and (CAG)n/CTG)n.

Shown below is a native 5-20% polyacrylamide gradient gel of the (CAG)n and (CTG)n, oligomers with the (CAG)n/(CTG)n duplex. Each sample is analyzed in parallel with the corresponding complementary duplex. **Lanes 1, 8, and 15**: molecular size markers in 10-bp increments. The bright band located approximately 1/3 from the top of the gel corresponds to the 100-bp marker. ; **Lanes 2, 3, and 4:** CAG4, CTG4, CAG4/CTG4; **Lanes 5, 6 and 7:** CAG10, CTG10, CAG10/CTG10; **Lanes 9,10, and 11:** CAG15, CTG15, CAG15/CTG15; **Lanes 12, 13, and 14:** CAG25, CTG25, CAG25/CTG25. NOTE: The faint band intensities for the CAG4 and CTG4 oligomers were lost due to the limitations of the data capture process, as well as the file conversion from a '.TIF file format to a Word document. In the original gel there was a single, but faint, band in each of the CAG4 and CTG4 lanes migrating just above the 10 kD marker.

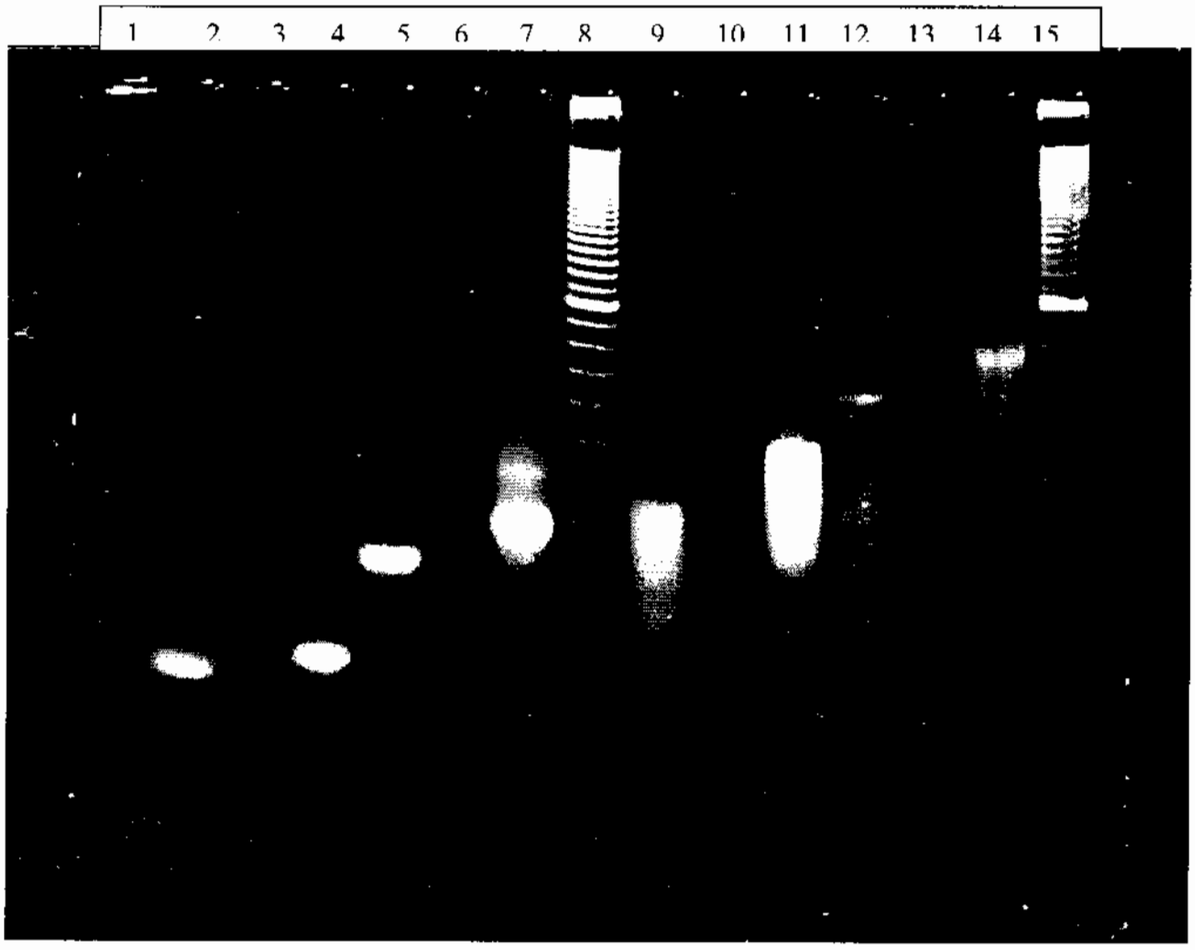


position corresponding to 45 base-pairs. The 75-mers for CAG and CTG showed a ladder that is most likely due to degradation of the oligomer during the incubation period. It is also quite possible that these sequences are forming slip-structures where the ideal blunt-ended complementarity is moved by one repeat unit. Under such conditions, the length and hence size, of the oliognucleotide changes in units of three depending upon the size of the slip. Nevertheless, there were bands appearing at the upper-most "rung" of the ladder at approximately 65 base-pairs for both CAG25 and CTG25. The complementary duplex appeared at approximately 75 to 77 base-pairs.

Analysis of the (CCG)n and (CGG)n oligomers showed solitary bands in each lane with a higher mobility than the corresponding duplex (Figure 61). The samples corresponding to the CCG4 and CGG4 oligomers migrated with an approximate size of 13 and 14 bp respectively, whereas the complementary duplex migrated with an apparent size of 16-17 bp. The bands corresponding to the CCG10 and CGG10 oligomers migrated with a position on the gel that corresponded to approximately 25 and 24 base-pairs, respectively, whereas the complementary duplex appeared at approximately 31 base-pairs. The CCG15 and CGG15 oligomers produced bands with an apparent mobility corresponding to 36 and 37 basepairs, respectively, while the complementary duplex migrated at a position corresponding to 48 base-pairs. The CCG25 and CGG25 oligomers each showed one band, migrating at approximately 60 and 61 base-pairs, respectively, in size. The complementary duplex migrated with a mobility corresponding to approximately 75 to 77 base-pairs.

Figure 61: Native gel electrophoresis of (CCG)n, (CGG)n and (CCG)n/CGG)n.

Shown below is a native 5-20% polyacrylamide gradient gel of the (CCG)n and (CGG)n, oligomers with the (CCG)n/(CGG)n duplex. Each sample is analyzed in parallel with the corresponding complementary duplex. **Lanes 1, 8, and 15**: molecular size markers in 10-bp increments. The bright band located approximately 1/3 from the top of the gel corresponds to the 100-bp marker.; **Lanes 2, 3, and 4:** CCG4, CGG4, CCG4/CGG4; **Lanes 5, 6 and 7:** CCG10, CGG10, CCG10/CGG10; Lanes 9,10, and 11: CCG15, CGG15, CCG15/CGG15; **Lanes 12, 13, and 14:** CCG25, CGG25, CCG25/CGG25.



While the observed size, in base-pairs, of the complementary duplexes corresponded to the anticipated size for each oligomer length under study, the observed size of the solitary oligomers, in basepairs of a fold-back hairpin, migrated with an apparently slower mobility. For example, an oligomer with 75 nucleotides, should fold back onto itself, forming a 36 base-pair hairpin (3 nucleotides in loop) that approximates a 36 base-pair duplex. It was therefore anticipated that a CNG25 oligomer would migrate at a position roughly corresponding to a 36 base-pair duplex. The fact all CNG25 oligomers migrated slower that the corresponding 36-mer suggests that either the structure is larger than predicted, due to a larger than assumed hairpin loop in concert with end-fraying, or that the charge on the solitary oligomer has been reduced, by counterion shielding of the phosphate ions along the DNA backbone. In addition, the mismatches along the hairpin could deform the DNA backbone by the formation of "bulges" within the mismatch region. These distortions can increase the apparent size of the hairpin, and produce the slower mobility pattern observed. Nevertheless, the gel electrophoretic data shows that the solitary oligomers have faster mobilities than the complementary duplex. Were these strands forming mismatched duplexes, the charge to size ratio would be similar to that observed for the complementary duplexes and, as a result, there would have been a band in the solitary oligomer lane that either comigrated with the complementary duplex or had a slightly slower mobility. In no example was such a band found, suggesting that the association is unimolecular, and that the structures formed by these oligomers are hairpin in nature. The gel electrophoresis data also suggests, from the lack of a ladder for all but the CAG25 and CTG25 sequences that the hairpins do not spontaneously form a

population of N-1 or N+1 slip-structures, but rather form the most thermodynamically stable "N" structure.

4.2.7 Differential Ion Binding Term

The differential ion binding term was measured for all (CNG)n oligomers, where N = A, C, T, or G and n = 4, 10 or 15 in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH 7.0 with sodium chloride added from a stock solution to produce a final concentration of 15, 25, 50, 100, 250, and 500 mM NaCl. The absorbance versus temperature data was collected then processed to obtain the melting temperature and the van't Hoff enthalpy for each oligomer at the specified salt concentration. The slope of a T_m versus log [Na^+] plot (Figure 62) was obtained from the linear regression fit (Table 18) and subsequently used to calculate the differential ion binding term using the equation previously described (Equation 24).

The data show several trends, beginning first with the magnitudes of the differential ion binding term, "n". The derived values of "n" for most oligomers, with the exception of the CGG and CTG oligomers, were below a numeric value of one suggesting that only one mole of sodium ion is released upon melting one mole of DNA strand. While examination of the data set did not reveal systematic errors, the data is unusual in many respects. In a more general sense, the trend of the data set paralleled the observations made with the circular dichroism spectroscopy. In particular, those oligomers found to have the highest level of red-shifting in the peak maxima at approximately 275 nm were also the same oligomers that have the higher differential ion binding terms. Of the three similar CD spectra, the CTG series of oligomers had the greatest red-shift in the position

Figure 62: Differential ion binding term plots for (CNG)n oligomers.

The plots used to derive the differential ion binding term are shown below for all oligomers examined. The slopes of the lines are used to determine the differential ion binding term.

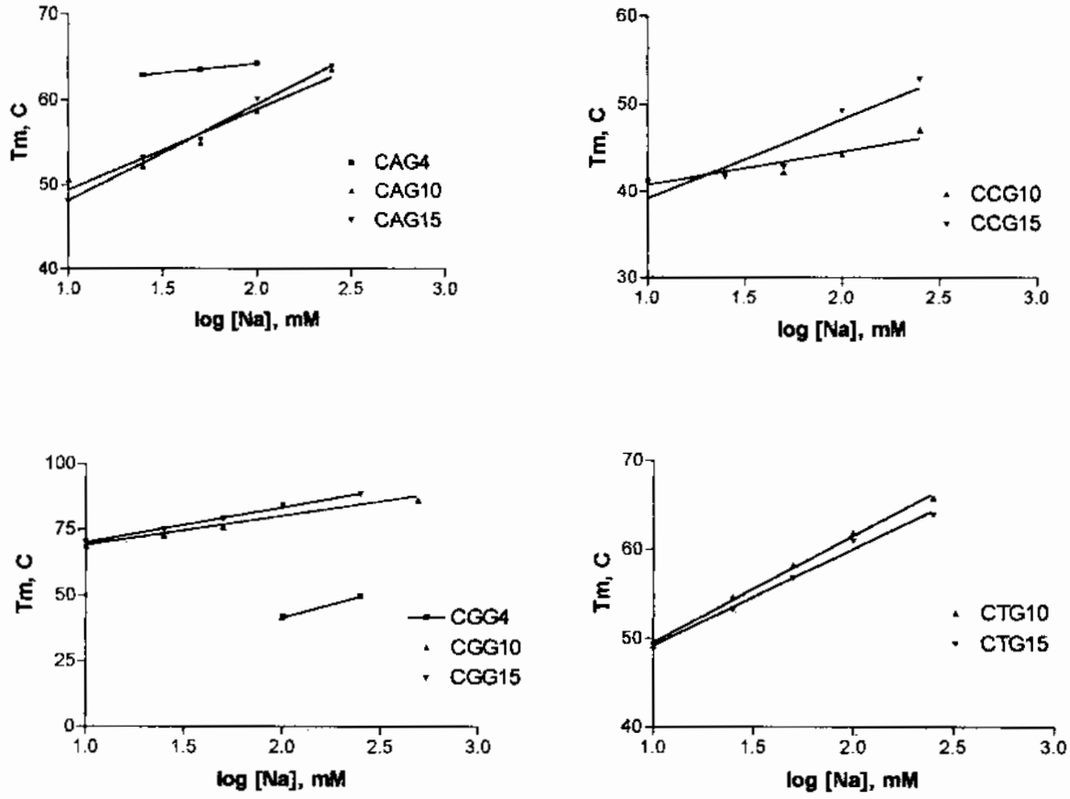


Table 18: Linear regression fit parameters of the differential ion binding term.

The linear regression parameters used to obtain the slope of the T_m versus log[Na^+] plot. The van't Hoff enthalpy, in kcal/mol, is for the single strand to hairpin transtion. All enthalpy measurements use the 100 mM sodium chloride sample.

Oligomer	Slope	Intercept	r^2	ΔH_{vh}	T_m (K)	dn
CAG4	2.4	60	0.99	-25134	337	0.12
CAG10	9.5	40	0.96	-26291	332	0.50
CAG15	11.4	37	0.99	-29002	333	0.65
CCG4	no transition					
CCG10	3.8	37	0.84	-28737	317	0.24
CCG15	9.1	30	0.88	-22484	322	0.43
CGG4	20.2	1.7	0.89	-20352	315	0.91
CGG10	11.1	11	0.89	-81205	357	1.54
CGG15	13.3	13	0.99	-65020	357	1.48
CTG4	no transition					
CTG10	12	12	0.99	-53257	335	1.24
CTG15	11	11	0.99	-47095	334	1.00

of the peak maxima (286 nm) than did the corresponding peak of oligomers CCG or CAG, with peak maximas located at 284 nm and 275 nm respectively (Figure 40, Figure 41). The red-shifted spectra of (CTG)n, suggested a lack of hydration that could possibly be the result of an increase in the uptake of sodium ions, as suggested by the higher differential ion binding term. The increased condensation of sodium ions not only allows the DNA to assume a more compact structure, but also decreases the base stacking interactions through realignment of the bases resulting from reorganization of the sugar-phosphate backbone.

4.3 Thermodyamics

4.3.1 Thermodynamic Parameters by UV-melting techniques

4.3.1.1 Thermodynamic Parameters as a Function of [Sodium Ion]

The differential ion binding term was measured for all (CNG)n oligomers, where N = A, C, T, or G and n = 4, 10 or 15 in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH 7.0 with sodium chloride added from a stock solution to produce a final concentration of 15, 25, 50, 100, 250, and 500 mM NaCl. The absorbance versus temperature data was collected then processed to obtain the melting temperature and the van't Hoff enthalpy for each oligomer at the specified salt concentration.

The (CAG)n oligomers studied show thermodynamic trends that paralleled the observations made within the circular dichroism studies (Table 19, Figure 55). The van't Hoff enthalpy for the single-strand (SS) to hairpin (HP) transition of CAG4 decreased

Table 19: Comparison of (CAG)n thermodynamic parameters versus [Na^+].

The van't Hoff thermodynamic parameters of the (CAG)4 – (CAG)15 oligomers as a function of Na^+ concentration are listed below. For all thermodynamic parameters listed below the assumed direction is the single-strand to hairpin transition.

Oligomer	[Na]	ΔH_{vh}	ΔG_{37C}	ΔS	T_m
	mM	kcal/mol	kcal/mol	cal/(mol K)	°C
CAG4	15	-25.04	-2.16	-73.80	66.23
CAG4	25	-23.98	-1.85	-71.39	62.91
CAG4	50	-24.01	-1.85	-71.51	63.55
CAG4	100	-25.13	-2.09	-74.34	64.26
CAG4	250	-27.60	-2.06	-82.38	61.93
CAG4	500	-18.00	-1.23	-54.10	59.42
CAG4	750	-22.15	-1.60	-66.30	60.15
CAG4	1000	-14.48	-0.68	-44.52	57.33

Oligomer	[Na]	ΔH_{vh}	ΔG_{37C}	ΔS	T_m
	mM	kcal/mol	kcal/mol	cal/(mol K)	°C
CAG10	15	-26.79	-1.14	-82.76	50.65
CAG10	25	-27.87	-1.31	-85.67	52.25
CAG10	50	-28.31	-1.55	-86.31	54.98
CAG10	100	-26.29	-1.72	-79.25	58.73
CAG10	250	-26.01	-2.05	-77.28	63.57
CAG10	500	-25.11	-1.97	-74.63	62.19
CAG10	750	-25.04	-1.88	-74.72	62.19
CAG10	1000	-22.26	-1.76	-66.10	63.97

Oligomer	[Na]	ΔH_{vh}	ΔG_{37C}	ΔS	T_m
	mM	kcal/mol	kcal/mol	cal/(mol K)	°C
CAG15	15	-30.10	-1.01	-93.82	48.02
CAG15	25	-29.05	-1.44	-89.04	53.18
CAG15	50	-30.86	-1.72	-94.00	55.33
CAG15	100	-29.00	-2.01	-87.07	60.10
CAG15	250	-31.40	-2.50	-93.23	63.89
CAG15	500	-29.68	-2.50	-87.67	65.55
CAG15	750	-29.63	-2.62	-87.15	67.00
CAG15	1000	-27.98	-2.34	-82.69	65.29

with increasing salt concentration. For CAG4, the height of the CD spectral peak at 275 nm remained constant until about 250 mM sodium chloride suggesting that part of the nucleotide is losing interaction beyond 250 mM Na^+. For CAG4, the van't Hoff enthalpy remained relatively constant, at approximately -25 kcal/mol until 500 mM Na^+ where the enthalpy decreased to approximately –18 kcal/mol. The entropy term for CAG4 also displayed similar trends. The entropy fluctuated between -70 and -80 cal/(mol K) up to approximately 500 mM Na^+ where the entropy became more favorable at -44 cal/(mol K). The increase in entropy is consistent with an oligomer that loses base stacking interactions due to an increase in the conformational flexibility of the DNA molecule. The trend of the stability, as indicated by the ΔG of the transition, suggests that the oligomers become less stable beyond 250 mM Na^+. This drop in stability is the result of the decrease in the enthalpic term that is not compensated by the increase in entropy. Samples of the CAG10 oligomer from 15 to 1000 mM Na^+ had shown a more subtle trend that mimicked the behavior of the 276 nm peak by circular dichroism. The height of the 276 nm peaks appeared to incrementally decrease from 15 to 1000 mM Na^+. The enthalpy for the SS to HP transition also decreased, but the increments were smaller than those observed for CAG4. The enthalpy at 15 and 25 mM Na^+ began at approximately –27 to –28 kcal/mol, and decreased by approximately –0.75 kcal per increase in sodium concentration. The entropy term for these samples became favorable, beginning from approximately -83 cal/(mol K) at 15 mM Na^+ to -66 cal/(mol K) at 1000 mM Na^+. Overall, the combined effects of the enthalpy and entropy changes increased the stability of the oligomer from 15 mM Na^+ to 250 mM Na^+. The stability, as indicated by the van't Hoff transition free energy, ΔG, reached a maximum at

approximately 250 mM Na^+. This is consistent with a DNA molecule that becomes stabilized with increasing salt concentration due to shielding of phosphate repulsion. In addition, the slight decrease in the stability and increase in entropy beyond 250 mM is consistent with a molecule that is losing hydration. It is interesting to note that a loss of hydration is not observed by circular dichroism, possibly because the hydration changes do not affect the nucleotide stacking. The CAG15 oligomers show trends that were similar, but did not completely mirror, the trends predicted by the CD spectra. The magnitude of the enthalpic term was between –29 and –31 kcal/mol throughout the range of 15 to 750 mM Na^+. The CD spectra predicted, on the basis of the observed decrease in the peak height at 275 nm, that the enthalpy term would begin to decrease, due possibly to the loss of interactions, at 250 mM Na^+. While the enthalpy did show a slight trend towards this direction beginning at 250 mM Na^+, the variation in the data was too large to make this determination with any certainty. The enthalpy decreased from –30 kcal/mol on average to –28 kcal/mol at 1000 mM Na^+. The entropy also displayed the same results, remaining relatively constant within the range of 15 to 250 mM Na^+. The entropy showed a marked increase at 1000 mM Na^+, - from -87 cal/(mol K) at 750 mM Na^+ to -83 cal/(mol K). The stability trend, as determined by the ΔG of the transition, did not reflect the changes observed in the enthalpy terms. The stability of the SS to HP transition became incrementally negative (more stable) as the Na^+, concentration increased to 750 mM Na^+ suggesting that the increase in entropy may drive the thermodynamic stability as a function of length.

The (CCG)n oligomers (Table 20, Figure 56) displayed thermodynamic trends that supported the observations made with the CD spectra. It is unfortunate, however, that the CCG4 oligomers displayed weak transitions, and could not be analyzed with existing UV-melting techniques. This suggests that the change in structure was accompanied by a lack of hyperchromicity. The CCG10 oligomers had shown an incremental decrease in enthalpy and a corresponding increase in entropy as the Na^+ concentration was increased. The enthalpy incrementally decreased from –51 kcal/mol at 15 mM Na^+ to –17 kcal/mol at 1000 mM Na^+. The CD spectra for this oligomer had shown a similar trend with respect to the peak at 285 nm. The height incrementally decreased with increasing salt concentration. The entropic term for CCG10 became more favorable, and increased from –163 cal/(mol K) at 15 mM Na^+ to –54 cal/(mol K) at 1000 mM Na^+. The stability of the hairpin, as reflected in the magnitude of the free energy for the SS to HP transition, remained relatively stable between 15 mM and 500 mM Na^+, but then became less negative with increased salt concentration. The enthalpy term for the CCG15 oligomers gradually became positive beginning from –53 kcal/mol, at 15 mM Na^+, to –17 kcal/mol at 750 mM Na^+. Like previously observed trends, the incremental decrease in enthalpy mirrors the incremental decrease in the observed circular dichroism peak at 284 nm. The prediction that the decrease in peak height is indicative of a decrease in interactions is supported by the enthalpic trends. It is possible that the peak at 284 nm indicates the relative amount of base stacking interactions for the nucleotides. The entropy term for the CCG15 oligomer became favorable, beginning from –168 cal/(mol K) at 15 mM Na^+ to –58 cal/(mol K) at 1000 mM Na^+. Unlike CCG10, the stability of the oligomer reached a maximum at approximately 250 mM Na^+, that was not reflected in either the

Table 20: Comparison of (CCG)n thermodynamic parameters versus [Na^+].

The van't Hoff thermodynamic parameters of the (CCG)4 – (CCG)15 oligomers as a function of Na^+ concentration are listed below. For all thermodynamic parameters listed below the assumed direction is the single-strand to hairpin transition. Those samples where no fit was possible, due to noise or lack of a transition, are denoted "no fit."

Oligomer	[Na]	ΔH_{vh}	ΔG_{37C}	ΔS	T_m
	mM	kcal/mol	kcal/mol	cal/(mol K)	°C
CCG4	15	no fit	no fit	no fit	no fit
CCG4	25	no fit	no fit	no fit	no fit
CCG4	50	no fit	no fit	no fit	no fit
CCG4	100	no fit	no fit	no fit	no fit
CCG4	250	no fit	no fit	no fit	no fit
CCG4	500	no fit	no fit	no fit	no fit
CCG4	750	no fit	no fit	no fit	no fit
CCG4	1000	no fit	no fit	no fit	no fit

Oligomer	[Na]	ΔH_{vh}	ΔG_{37C}	ΔS	T_m
	mM	kcal/mol	kcal/mol	cal/(mol K)	°C
CCG10	15	-51.46	-0.75	-163.60	41.57
CCG10	25	-44.84	-0.75	-142.24	42.28
CCG10	50	-43.09	-0.70	-136.76	42.18
CCG10	100	-28.74	-0.55	-91.52	44.26
CCG10	250	-22.48	-0.71	-70.24	47.08
CCG10	500	-20.55	-0.84	-63.59	50.08
CCG10	750	-17.47	-0.58	-54.47	47.82
CCG10	1000	no fit	no fit	no fit	no fit

Oligomer	[Na]	ΔH_{vh}	ΔG_{37C}	ΔS	T_m
	mM	kcal/mol	kcal/mol	cal/(mol K)	°C
CCG15	15	-53.00	-0.68	-168.78	41.00
CCG15	25	-46.40	-0.70	-147.41	41.76
CCG15	50	-42.26	-0.79	-133.75	42.90
CCG15	100	-30.00	-1.14	-93.11	49.22
CCG15	250	-27.93	-1.37	-85.67	52.87
CCG15	500	-21.60	-0.99	-66.48	51.96
CCG15	750	-21.21	-1.25	-70.86	54.04
CCG15	1000	-19.30	-1.09	-58.74	56.51

CD spectra or the trends observed for CCG10. The base stacking interactions, as reflected in the trend of the CCG15 CD spectra incrementally decreased as the Na^+ concentration increased. Therefore, any stability gains, as shown by the free energy trend must be due to an increase in the entropy term. Once again, loss of water and uptake of sodium ions increased the oligomer stability. At high salt, the process is entropically driven, as expected for a process driven by salt effects.

The (CGG)n oligomers had once again shown trends that were different than those observed for the CAG, CCG, or CTG oligomers (Table 21, Figure 57). Overall, the observed trends were consistent with the CD spectra. The CGG4 oligomers had, at low Na^+ concentration dual transitions between 15 and 50 mM Na^+. Dual transitions within this range are consistent with an oligomer that, as was indicated by the CD spectra, did not achieve the final structural form until 50 to 100 mM Na^+. The dual transitions are indicative of two different structural forms within the melting population, and suggest non-two-state interactions at low Na^+ concentrations. This is consistent with the CD spectra for CGG4, where, at 15 mM Na^+ the spectrum bore no resemblance to the spectra for other salt concentrations. As the salt concentration was increased to 25 mM, there was a structural transition, but the transition was located somewhere between the final structure, achieved at 100 mM Na^+, and the starting structure shown by 15 mM Na^+. Because of the dual transitions observed for CGG4 between 15 and 50 mM Na^+ the extraction of thermodynamic data was not attempted. The remainder of the CGG4 series, from 100 mM Na^+ to 1000 mM Na^+ revealed that the enthalpy increased from –20 kcal/mol to –31 kcal/mol. Consistent with enthalpy/entropy compensation, the entropic

Table 21: Comparison of (CGG)n thermodynamic parameters versus [Na^+].

The van't Hoff thermodynamic parameters of the (CGG)4 – (CGG)15 oligomers as a function of Na^+ concentration are listed below. For all thermodynamic parameters listed below the assumed direction is the single-strand to hairpin transition. Samples of CGG4 where no fit was possible, due to noise or lack of a transition, are denoted "no fit." The thermodynamic parameters for samples with a superscript of one were determined by first derivative and may differ from the parameters determined by van't Hoff analysis.

Oligomer	[Na]	ΔH_{vh}	ΔG_{37C}	ΔS	T_m
	mM	kcal/mol	kcal/mol	cal/(mol K)	°C
CGG4	15	dual trans	dual trans	dual trans	dual trans
CGG4	25	dual trans	dual trans	dual trans	dual trans
CGG4	50	dual trans	dual trans	dual trans	dual trans
CGG4	100	-20.35	-0.49	-64.07	41.62
CGG4	250	-24.10	-0.92	-74.77	49.49
CGG4	500	-27.31	-1.24	-84.09	51.62
CGG4	750	-29.77	-1.46	-91.33	52.94
CGG4	1000	-31.52	-1.61	-96.49	53.69

Oligomer	[Na]	ΔH_{vh}	ΔG_{37C}	ΔS	T_m
	mM	kcal/mol	kcal/mol	cal/(mol K)	°C
CGG10	15	-54.07	-5.02	-158.22	68.73
CGG10	25	-56.69	-5.85	-164.00	72.65
CGG10	50	-52.98	-5.89	-151.89	75.76
CGG10	100	no fit	no fit	no fit	no fit
CGG10	250	-57.92	-6.48	-165.95	76.49
CGG10	500	-83.93	-11.57	-233.42	86.55
CGG10	750	-91.08	-12.72	-252.75	87.34
CGG10	1000	-87.52	-12.24	-242.83	87.57

Oligomer	[Na]	ΔH_{vh}	ΔG_{37C}	ΔS	T_m
	mM	kcal/mol	kcal/mol	cal/(mol K)	°C
CGG15	15	-72.60	-7.04	-211.48	70.28
CGG15	25	-71.65	-7.81	-205.93	74.92
CGG15	50	-71.20	-8.50	-202.27	79.01
CGG15	100	-65.02	-8.55	-182.17	83.94
CGG15	250[1]	-13.58	-1.92	-37.60	88.45
CGG15	500[1]	-11.61	-1.69	-32.00	89.72
CGG15	750[1]	-12.84	-1.88	-35.36	90.11
CGG15	1000[1]	-11.44	-1.71	-31.38	91.46

contribution for this range also became less favorable, from –64 cal/(mol K) to –96 cal/(mol K). In spite of these changes the stability of the hairpin increased from –0.5 kcal/mol at 100 mM Na^+ to –1.6 kcal/mol at 1000 mM Na^+. In this case, the free energy of stability was not entropically driven, as was observed for many of the other oligomers, but was enthalpically driven. The increase in enthalpy for the CGG4 oligomer is mirrored by the parallel shift in the depth of the trough at 257 nm and the peak at 281 nm, as well as by the increased peak height at 240 nm. Together with the CD data, these results suggest that the oligomer undergoes an increase in base stacking interactions, with a decrease in entropy, due to a decrease in conformational flexibility. Were the loss of entropy due to a hydration increase, the CD spectra would have shown an overall shift towards the blue end of the spectrum, which was not observed. The thermodynamic results for the CGG10 oligomers produced the same trends as observed for the CGG4 oligomers. The enthalpy increased from –54 kcal/mol at 15 mM Na^+ to –87 kcal/mol at 1000 mM Na^+. The entropy became less favorable, decreasing from –158 cal/(mol K) at 15 mM Na^+ to –242 cal/(mol K) at 1000 mM Na^+. This is consistent with the CD spectra that suggested the base stacking interactions, as indicated by the trough/peak at 257/281increased as a function of Na^+ concentration. The height of the peak at 240 nm increased and the depth of the through at 222 nm became more negative in a manner consistent with these conclusions. The stability of the oligomer, indicated by the free energy of the SS to HP transition, increased from –5 kcal/mol at 15 mM Na^+ to –12 kcal/mol at 1000 mM Na^+. This suggests, in consideration of the unfavorable entropy term, that the hydration process for CGG10 is enthalpically driven. The CD spectra for the CGG15 samples did not mirror the thermodynamic trends observed for the CGG10 or

CGG4 samples (Table 21). The enthalpy decreased from –73 kcal/mol at 15 mM Na^+ to –11 kcal/mol at 1000 mM Na^+. The trend of the CD spectra for this sample suggested however, that the base stacking interactions, and hence the enthalpy should have increased with increasing salt concentrations (Figure 57). Examination of the sample set revealed that the thermodynamic data for the latter part of the CGG15 samples were reduced using a first derivative, and not a van't Hoff model. A first derivative model was necessary because of the noise level and high melting temperatures for the samples under consideration. It is therefore possible, given the conspicuous change in the enthalpic term that begins with the utilization of the new model, that the enthalpy and free energy terms for all CGG15 samples from 250 mM Na^+ to 1000 mM Na^+ have an abnormally low value to an inherent experimental bias. The stability trend, reflected the enthalpy trend, and showed that the oligomer became less stable with increasing salt concentration. The free energy decreases from –7 kcal/mol at 15 mM Na^+ to –1.7 kcal/mol at 1000 mM Na^+. This trend is in direct opposition of the trend of the melting temperature that suggests the CGG15 oligomer is becoming more stable with increasing Na^+ concentration. Again , it is very likely that this effect is due to the utilization of two different models for the data set rather than some underlying biophysical phenomenon. The overall driving "force" for this process however, is still enthalpy, as was observed for the shorter oligomers.

The CTG oligomers, were similar in thermodynamic behavior to the (CAG)n and (CCG)n oligomers (Table 22, Figure 58). The CTG4 oligomers displayed a lack of hyperchromicity coupled with weak transitions that could not be readily fit to van't Hoff

Table 22: Comparison of (CTG)n thermodynamic parameters versus [Na^+].

The van't Hoff thermodynamic parameters of the (CTG)4 – (CTG)15 oligomers as a function of Na^+ concentration are listed below. For all thermodynamic parameters listed below the assumed direction is the single-strand to hairpin transition. Those samples where no fit was possible, due to noise or lack of a transition, are denoted "no fit."

Oligomer	[Na]	ΔH_{vh}	ΔG_{37C}	ΔS	T_m
	mM	kcal/mol	kcal/mol	cal/(mol K)	°C
CTG4	15	no fit	no fit	no fit	no fit
CTG4	25	no fit	no fit	no fit	no fit
CTG4	50	no fit	no fit	no fit	no fit
CTG4	100	no fit	no fit	no fit	no fit
CTG4	250	no fit	no fit	no fit	no fit
CTG4	500	no fit	no fit	no fit	no fit
CTG4	750	no fit	no fit	no fit	no fit
CTG4	1000	no fit	no fit	no fit	no fit

Oligomer	[Na]	ΔH_{vh}	ΔG_{37C}	ΔS	T_m
	mM	kcal/mol	kcal/mol	cal/(mol K)	°C
CTG10	15	-50.60	-1.90	-157.08	49.13
CTG10	25	-49.00	-2.63	-149.57	54.59
CTG10	50	-49.03	-3.13	-148.06	58.19
CTG10	100	-53.26	-3.96	-159.03	61.87
CTG10	250	-47.57	-4.05	-140.42	65.80
CTG10	500	-47.01	-4.29	-137.82	68.06
CTG10	750	-45.74	-4.24	-133.88	68.65
CTG10	1000	-43.00	-4.05	-125.65	69.24

Oligomer	[Na]	ΔH_{vh}	ΔG_{37C}	ΔS	T_m
	mM	kcal/mol	kcal/mol	cal/(mol K)	°C
CTG15	15	-49.86	-1.89	-154.72	49.23
CTG15	25	-48.09	-2.39	-147.42	53.23
CTG15	50	-50.37	-3.01	-152.78	56.73
CTG15	100	-47.10	-3.38	-141.03	60.94
CTG15	250	-45.34	-3.79	-134.04	63.85
CTG15	500	-47.83	-4.40	-140.11	67.64
CTG15	750	-46.83	-4.44	-136.73	69.39
CTG15	1000	-42.43	-3.95	-124.13	68.82

analysis. The enthalpy of the CTG 10 oligomers decreased form –50 kcal/mol in 15 mM Na^+ to –43 kcal/mol in 1000 mM Na^+. This was consistent with the behavior of the CD spectra, where the height of the peak observed at 285 nm incrementally decreased from 15 to 500 mM Na^+. The entropy for this same range in Na^+ concentration increased from –157 cal/(mol K) to –125 cal/(mol K). The stability of this oligomer increased from –1.9 kcal/mol to –4 kcal/mol. Overall the observed trends suggest that the hydration process for CTG10 is entropically driven.

Overall, the enthalpy of CTG15 oligomers decreases from –49 kcal/mol in 15 mM Na^+ to –42 kcal/mol in 1000 mM Na^+. The point-to-point trend of the enthalpy data mimics the behavior observed for the CD spectra of CTG15. Namely, the enthalpy incrementally decreased from approximately –50 kcal/mol in 25 mM Na^+, to about –45 kcal/mol in 250 mM Na^+. Like the observed trend in the CD spectra, the enthalpy increased at 500 mM Na^+ to –47 kcal/mol, then incrementally decreased to –42 kcal/mol in 1000 mM. Overall, the entropy for this same range in Na^+ concentration increased from –154 cal/(mol K) to –124 cal/(mol K). Like the enthalpy term the pattern, but not the direction, of the entropy mirrored the behavior observed in the CD spectra. For each decrease in enthalpy, there was a corresponding increase in entropy. The entropy incrementally increased from approximately –150 cal/(mol K) in 25 mM Na^+, to about –134 cal/(mol K) in 250 mM Na^+. The enthalpy decreased at 500 mM Na^+ to –140 kcal/mol, then incrementally increased to –124 cal/(mol K) in 1000 mM. In spite of the wild fluctuations observed for the enthalpy and entropy terms, the stability of this oligomer

increased from – 1.9 kcal/mol to –3.9 kcal/mol and suggests that the hydration of CTG15, like that observed for CAG 15, is entropically driven.

4.3.1.2 Comparison of Oligomer and Duplex van't Hoff Stabilities (Oligomer Concentration Adjusted).

The thermodynamic parameters for the (CNG) triplet repeat sequences were initially determined by optical melting techniques in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH7.0, 100 mM NaCl, with DNA concentrations between 1.0 μM and 9.5 μM. The absorbance versus temperature data (Figure 63, Figure 64, Figure 65, Figure 66 , Figure 67, Figure 68) were collected and processed as previously described (3.7.2) to obtain the transition melting temperature, the van't Hoff enthalpy, entropy, and Gibbs free energy for both the hairpin to single-strand transition and the duplex to single-strand transition. For all data sets the assumed directionality of the thermodynamic parameters was the SS to HP direction. Therefore, increased stability was achieved with increased negativity of the free energy, increased interaction with increased negativity of the enthalpy, and increased disorder with increased positivity of the entropy.

The (CAG)n oligomers, where n = 4, 10, 15, or 25 were examined with respect to increasing length (Figure 63, Table 23). The enthalpy of the hairpin (HP) to single-strand (SS) transition became incrementally more favorable as a function of strand length. By the same measure, the entropy of this transition became less favorable as function of increasing length. Between the lengths of 12 and 45 nucleotides, the free energy fluctuated around -2 kcal/mol, but then finally became more negative, -2.7 kcal/mol, at

Figure 63: Representative UV-thermal melts of the (CAG)n oligomers.

Selected absorbance versus temperature melt data for the (CAG)n oligomers in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH7.0, 100 mM NaCl are shown below.

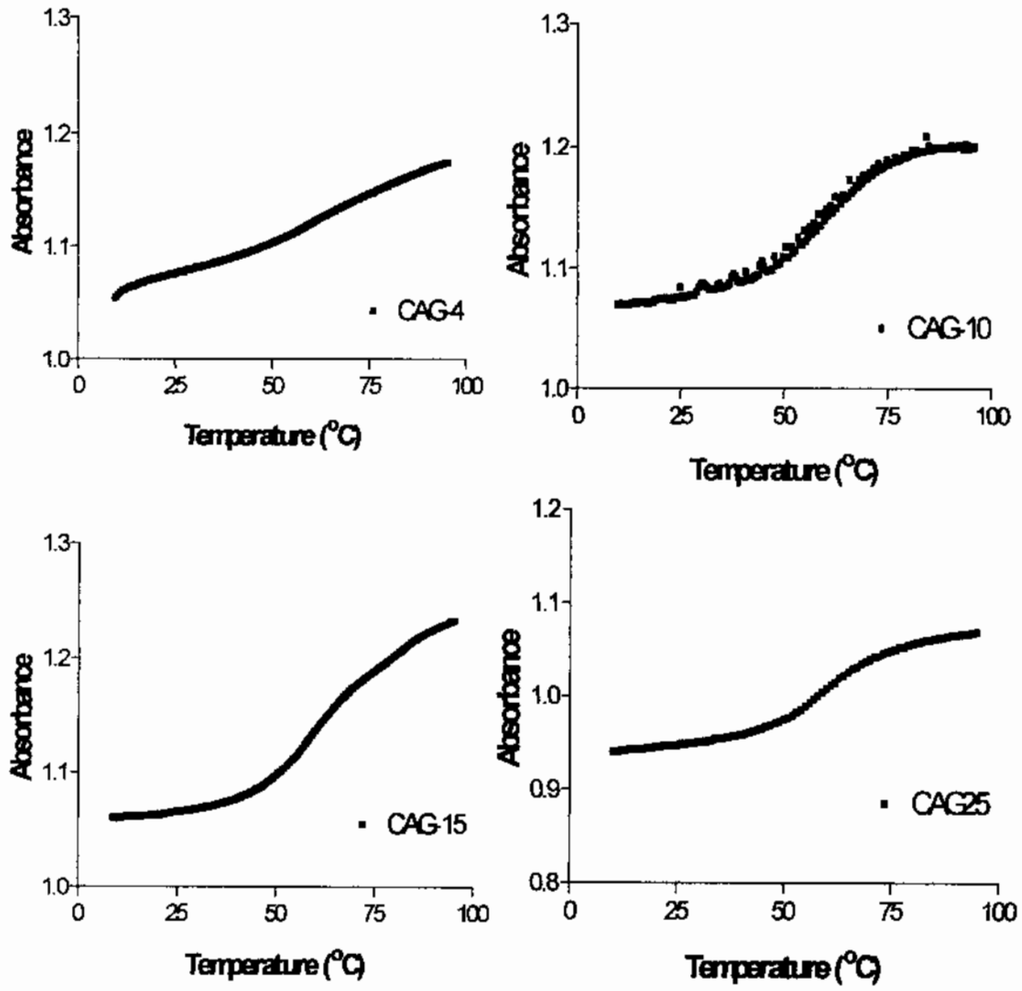


Figure 64: Representative UV-thermal melts of the (CCG)n oligomers.

Selected absorbance versus temperature melt data for the (CCG)n oligomers in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH7.0, 100 mM NaCl are shown below.

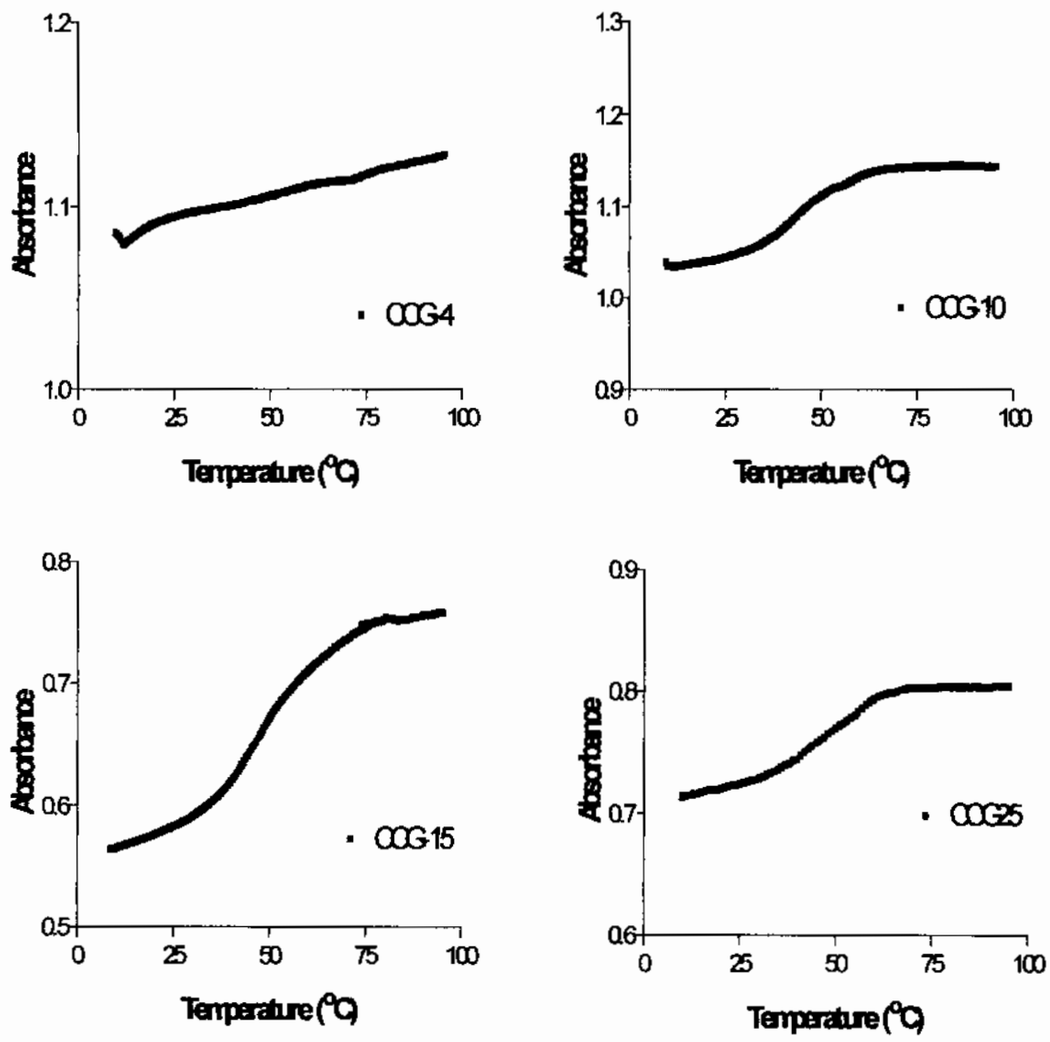


Figure 65: Representative UV-thermal melts of the (CGG)n oligomers.

Selected absorbance versus temperature melt data for the (CGG)n oligomers measured in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH7.0, 100 mM NaCl are shown below.

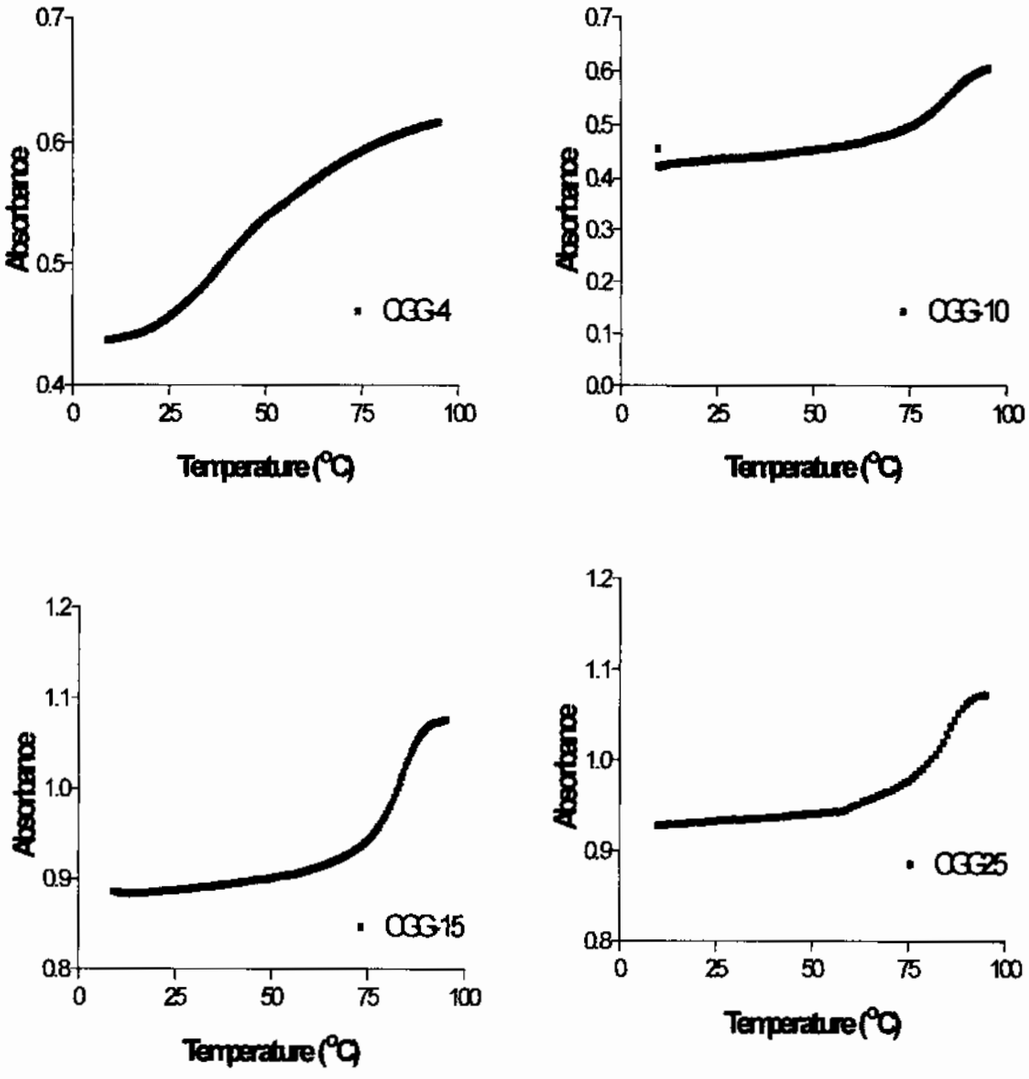


Figure 66: Representative UV-thermal melts of the (CTG)n oligomers.

Selected absorbance versus temperature melt data for the (CTG)n oligomers measured in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH7.0, 100 mM NaCl are shown below.

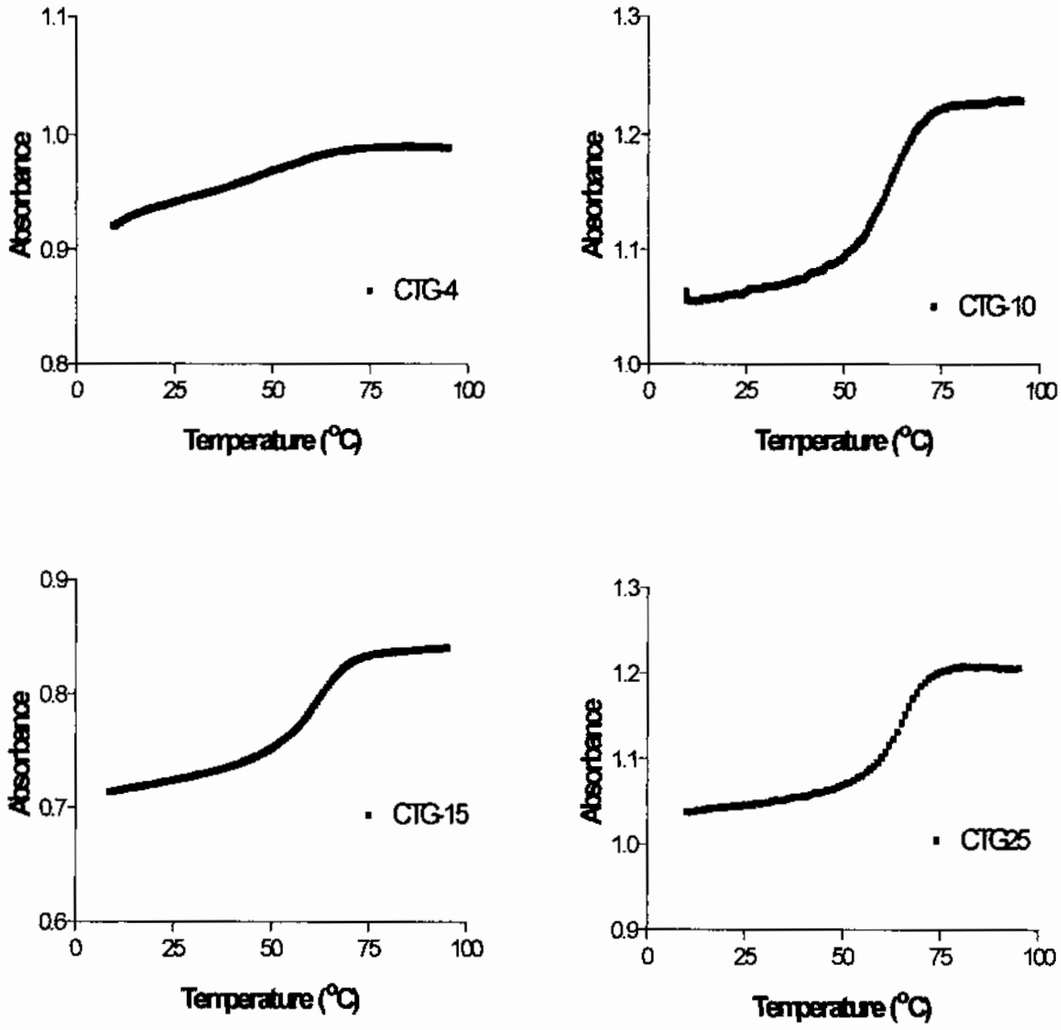


Figure 67: Representative UV-thermal melts of the (CAG)n/(CTG)n duplexes.

Selected absorbance versus temperature melt data for the (CAG)n/(CTG)n duplexes examined are shown below.

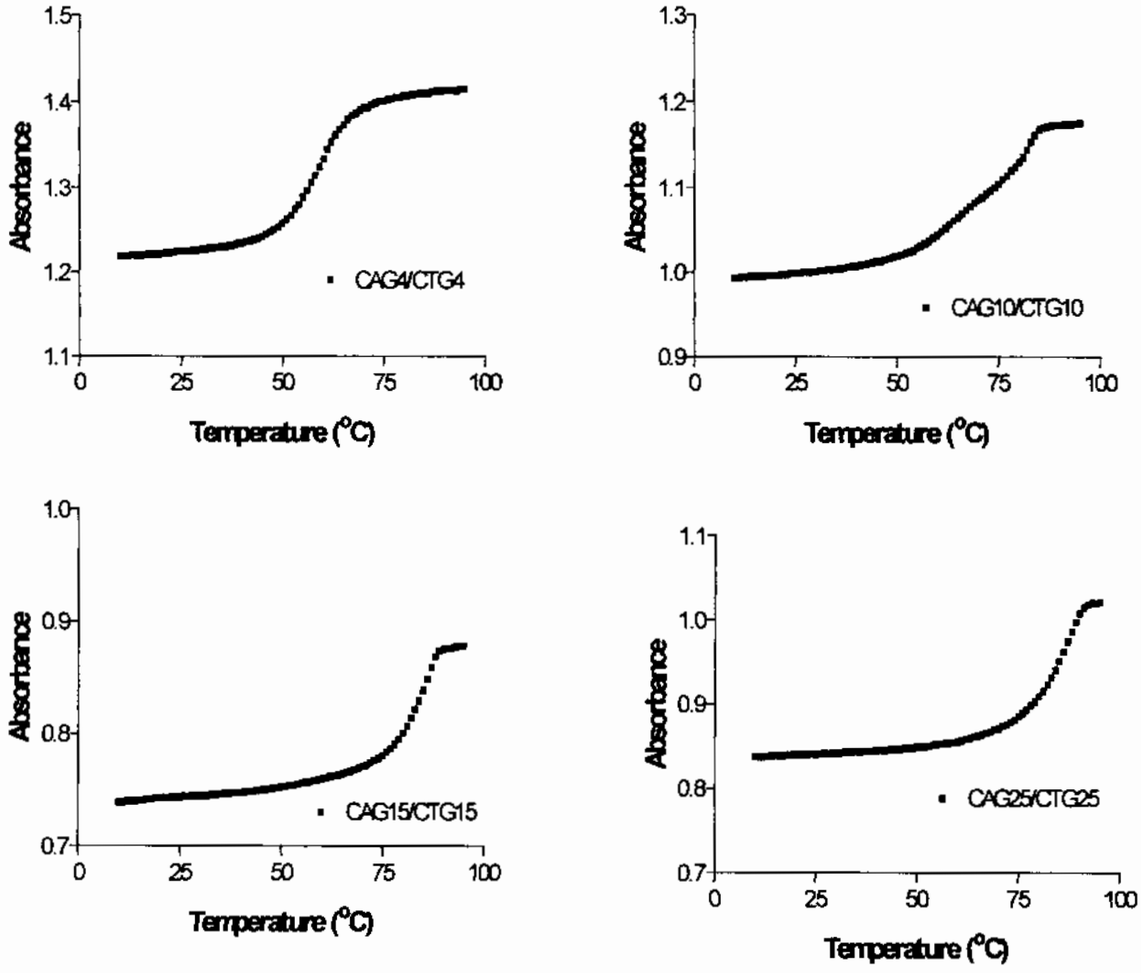


Figure 68: Representative UV-thermal melts of the (CCG)n/(CGG)n duplexes.

Selected absorbance versus temperature melt data for the (CCG)n/(CGG)n duplexes examined are shown below.

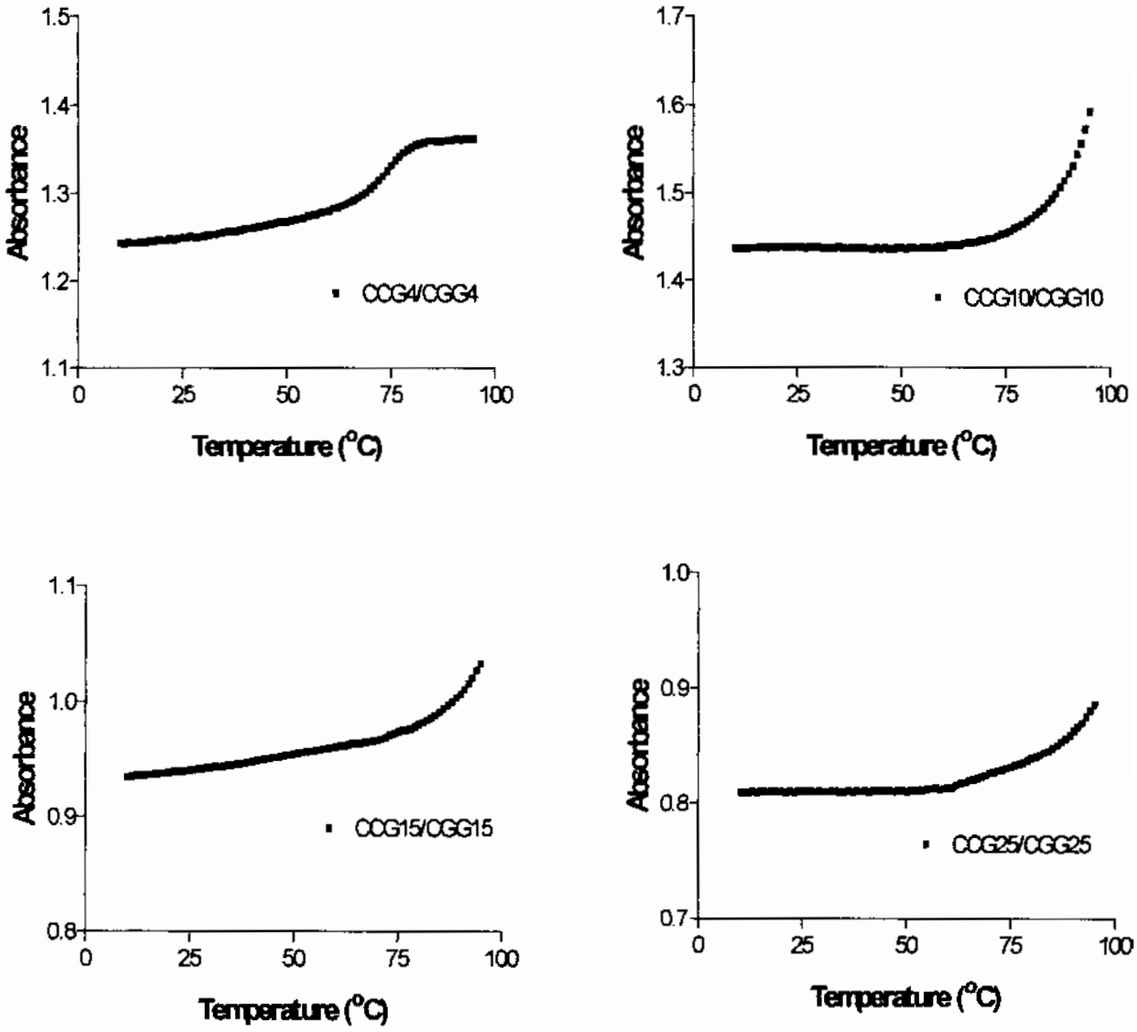


Table 23: Comparison of UV-melt thermodynamic parameters vs. HP length.

The van't Hoff thermodynamic parameters of the (CNG)n oligomers as a function of oligomer length are listed below. For all thermodynamic parameters the assumed direction is the single-strand to hairpin transition. Those samples where no fit was possible, due to noise or lack of a transition, are denoted "no fit."

Oligomer	[Na]	ΔH_{vh}	ΔG_{37C}	ΔS	T_m
	mM	kcal/mol	kcal/mol	cal/(mol K)	°C
CAG4	100	-25.1	-2.1	-74.3	64.26
CAG10	100	-26.3	-1.7	-79.2	58.73
CAG15	100	-29.0	-2.0	-87.1	60.10
CAG25	100	-40.6	-2.7	-122.3	58.79
Oligomer	**[Na]**	**ΔH_{vh}**	**ΔG_{37C}**	**ΔS**	**T_m**
	mM	kcal/mol	kcal/mol	cal/(mol K)	°C
CCG4	100	no fit	no fit	no fit	no fit
CCG10	100	-28.7	-0.5	-91.5	44.26
CCG15	100	-30.0	-1.1	-93.1	49.22
CCG25	100	-27.5	-1.0	-85.2	48.93
Oligomer	**[Na]**	**ΔH_{vh}**	**ΔG_{37C}**	**ΔS**	**T_m**
	mM	kcal/mol	kcal/mol	cal/(mol K)	°C
CGG4	100	-20.4	-0.5	-64.1	41.62
CGG10	100	no fit	no fit	no fit	no fit
CGG15	100	-65.0	-8.5	-182.2	83.94
CGG25	100	-27.7	-3.5	-78.0	81.67
Oligomer	**[Na]**	**ΔH_{vh}**	**ΔG_{37C}**	**ΔS**	**T_m**
	mM	kcal/mol	kcal/mol	cal/(mol K)	°C
CTG4	100	no fit	no fit	no fit	no fit
CTG10	100	-53.3	-4.0	-159.0	61.87
CTG15	100	-47.1	-3.4	-141.0	60.94
CTG25	100	-41.4	-3.4	-122.8	64.49

75 nucleotides. These trends indicate that the stability of (CAG)n series of oligomers is enthalpically driven. While the entropy compensates for the gains made by the enthalpy term, the overall deciding factor was the enthalpic contribution to the overall stability. This indicates that the stability of the CAG hairpins comes from increased hydrogen bonding and/or base stacking interactions as the length of the DNA increases. As expected, the increased interaction serves to reduce the conformational flexibility of the oligomer making the entropy less favorable. In spite of these trends, the gains made by the enthalpic term to the stability of these sequences is, at best modest, and amounted to approximately 1 kcal/mol of stability overall. This predicts that the CAG oligomers will be among the least stable of all sequences studied as a function of increasing length. When contrasted with the CD spectral results, adjusted for oligomer concentration, the incremental increase in the enthalpy mirrors the incremental increase observed in the peak at 275 nm, and the decrease in the trough at 251 nm (Figure 40, A). This observation suggests that the increases in enthalpy are related to reorientation of the bases resulting in increased base stacking interactions for the entire oligomer. The (CCG)n oligomers, where n = 10, 15, or 25, produced trends that were not easy to discern due to the narrow fluctuations in the magnitudes of the enthalpy, entropy and free energy terms for all lengths studied. The oligomer CCG4 produced weak transitions and made determination of thermodynamic parameters based on this method questionable, at best. The free energy of the CCG10 through CCG25 oligomers began from –0.5 kcal/mol, became more favorable by approximately 0.6 kcal/mol and reached a final value of approximately -1 kcal/mol at with CCG25. In this same range of nucleotides, the enthalpy became less favorable, from –29 kcal/mol to approximately –27 kcal/mol, while

the entropy became more favorable from –91 cal/(mol K) to –85 cal/(mol K). The data suggest that the modest increase in stability observed is a result of an increase in the entropy as a function of length, in conjunction with a loss of hydrogen-bond formation or base stacking interactions. These thermodynamic results are in contrast to the CD spectral results (adjusted for oligomer concentration) that showed an increase in the height of the peak at 285 nm, and an increase in the depth of the trough at 256 nm with increasing length (Figure 40, B). The CD data suggest that the apparent loss in enthalpy may be due to decreased H-bond formation, rather than a loss base stacking interactions. Such a scenario is possible if there is significant end-fraying as a function of length, which would serve to raise the entropy of the molecule, and decrease the overall number of H-bonds, with a simultaneous reordering of the mismatched nucleotides, to achieve the greater base stacking interactions within the stem region of the hairpin. The overall net effect is that the stability of these oligomers appeared to reach a plateau, both in terms of enthalpy and in free energy, with respect to length. Like CAG, the oligomer does not significantly become more stable as a function of length. The (CTG)n oligomers where n = 10, 15 and 25, displayed clear trends that paralleled the observations made with the CD spectra (Figure 41, B). Unlike the CD spectral data, however, derivation of thermodynamic parameters from the CTG4 melt data was difficult for CTG4 due to the shallow nature of the transition, and left open the possibility of a non two-state interaction. The free energy of these oligomers, over the range studied, slightly decreased from –4.0 kcal/mol at 30 nucleotides to –3.4 kcal/mol at 75 nucleotides. Within this same range, the entropy became more favorable, from –160 cal/(mol K) to –120 cal/(mol K), while the enthalpy became less favorable, from –53 cal/mol with

CTG10 to –41 kcal/mol with CTG25. The data suggests that the stability of CTG oligomers is enthalpically driven, but the enthalpy and hence the free energy decreases as a function of length. There is, for (CTG)n an enthalpically driven instability. In addition, the free energy, while decreasing from 30 to 75 nucleotides, appeared to plateau by 45 nucleotides. Like the CCG oligomers, the observed decrease in enthalpy, the observed increase in entropy, and the increase in the peak height at 286 nm suggest that there is increased interaction within the stem region of the hairpin, possibly due to reorientation of the mismatched nucleotides. This increase in interaction however, is not enough to compensate for the loss of hydrogen bond formation due possibly to end-fraying near the terminus of the oligonucleotide, or expansion within the loop region of the oligomer. In spite of the fact that the entropy becomes larger, the gains in conformational flexibility are not enough to offset the loss of bond formation.

The (CGG)n oligomers, where n = 4, 10, 15 or 25 show unusual trends that have become the hallmark of this type of DNA. Across all therrmodynamic parameters there was a peak in the trends that occurred at approximately 45 nucleotides. The enthalpy becomes more favorable from –20 kcal/mol with CGG4, to –65 kcal/mol with CGG15. The enthalpy then decreases to –27 kcal/mol by 75 nucleotides. The entropy displayed similar trends, beginning from –64 cal/(mol K) and becoming unfavorable at 45 nucleotides in length with –180 cal/(mol K). The entropy became favorable at 25 nucleotides in length and achieved a final magnitude of –78 cal/(mol K). The stability for these oligomers also produced matching trends. The stability reached a maximum at 45 nucleotides, then decreased from –8.5 kcal/mol at 45 nucleotides to –3.5 kcal/mol at 75

nucleotides. Overall, the data displayed enthalpy/entropy compensation, where the magnitude of the entropy became more favorable with loss of enthalpy. The data trend suggest that the stability of the CGG oligomers are enthalpically linked. The observed enthalpy trend suggests that there should also be a similar increase then decrease trend in the region of the CD spectra corresponding to the base stacking interactions of the oligomer. However, for most measurable parameters the height of the peaks at 282 and 240 nm, after concentration adjustment, suggest that the interactions increased with length (Figure 41, A). It is quite possible however that the observed thermodynamic trends do not reflect the underlying biophysics of the CGG oligomers, but rather the peculiarities of the data itself, and the data reduction process.

The duplexed oligomers produced several results worth noting. All oligomers were examined as equal length complementary duplexes so that a reference state for the hairpins could be measured. The (CAG)n/(CTG)n complementary duplexes produced thermograms that underwent a complete but asymmetric transition that was biphasic in nature. These thermograms were analyzed with van't Hoff analysis in an attempt to extract meaningful thermodynamic information from these melts, with the constraint that there could exist a bias in the data. The (CCG)n/(CGG)n complementary duplexes underwent transitions up to and including the melting transition. However, the nature of the data extraction and van't Hoff analysis process with the Cary Varian "Thermal" software precluded accurate determination of the melting transition, and the thermodynamic parameters. With respect to the CAG/CTG duplexes, the enthalpy became less favorable as function of length, beginning from –70 kcal/mol at 12 base-

pairs in length, and decreasing to –29 kcal/mol at 75 base-pairs in length. The entropy within this range became favorable, from –212 cal/(mol K) to –82 cal/(mol K). The stability of the oligomers appeared to decrease, from –4.4 kcal/mol to –3.8 kcal/mol as the length increased form 12 to 75 base-pairs. The observed trends suggest that the CAG/CTG duplexes decrease stability, as a function of decreasing enthalpy with respect to length. The observed stability trend – the decrease in enthalpy as a function of length – is in contrast with the observed CD spectral data (Figure 42, A). The CD spectra show relatively little change in the stacking interactions, as reflected in the relatively small change in the height of the peak at 277 nm. The stability trends are also in disagreement with the observed trends for the melting transition temperature for all other oligomers studied to date. Generally, as the stability of the oligomer increased, the temperature of the melting transition also increased. Comparison of the trend data therefore suggests that the observed derived thermodynamic parameters – the enthalpy, entropy and free energy – may contain a systematic bias that is depressing the data near the long-length end of the data set. Such sources of error were found to derive from the selection of the region of the alpha curve for van't Hoff analysis. Curves with two apparent slopes, such as was observed for these samples, could alter the derived enthalpy value of the longer duplexes, which had a more severe biphasic slope problem than the shorter duplexes.

In general, the thermodynamic trends also point to an unusual observation (Table 24, Table 25). In a general view, small hairpins are relatively stable, but less stable than longer oligomers that can form long hairpins. This fact was shown, in detail, within previous sections. However, on a nucleotide adjusted basis, the enthalpy per nucleotide

Table 24: Comparison of UV-melt thermodynamic parameters vs. DUP length.

The van't Hoff thermodynamic parameters of the (CNG)n/(CNG)n oligomers as a function of duplex length are listed below. For all thermodynamic parameters listed below the assumed direction is the single-strand to hairpin transition. Those samples with incomplete transitions, where the midpoint of the UV-melt curve could not be clearly defined, are labeled "inc. trans."

Oligomer	[Na]	ΔH_{vh}	ΔG_{37C}	ΔS	T_m
	mM	kcal/mol	kcal/mol	cal/(mol K)	°C
CAG4/CTG4	100	-70.3	-4.4	-212.5	57.84
CAG10/CTG10	100	-45.1	-4.4	-131.1	70.71
CAG15/CTG15	100	-41.4	-5.2	-116.6	81.80
CAG25/CTG25	100	-29.2	-3.8	-81.8	83.83
Oligomer	**[Na]**	ΔH_{vh}	ΔG_{37C}	ΔS	T_m
	mM	kcal/mol	kcal/mol	cal/(mol K)	°C
CCG4/CGG4	100	-27.9	-2.7	-81.2	70.03
CCG10/CGG10	100	inc. trans	inc. trans	inc. trans	inc. trans
CCG15/CGG15	100	inc. trans	inc. trans	inc. trans	inc. trans
CCG25/CGG25	100	inc. trans	inc. trans	inc. trans	inc. trans

Table 25: Comparison of the oligomer enthalpy per nucleotide.

A. The van't Hoff enthalpy, in kcal/mol nucleotide, derived from the UV-melt thermodynamic data, was divided by the length of the oligonucleotide to obtain the gain in enthalpy as a function of oligomer length. The data suggest that increasing the strand length does not increase the enthalpy in a linear fashion as length increases. The longer strands are therefore less stable than would be theoretically predicted if the total enthalpy were simply an additive sum of enthalpies for each triplet repeat unit. Enthalpy in kcal/mol. Enthalpy/nucleotide calculated in terms of nucleotides. **B**. The theoretical enthalpy in kcal/mol was calculated assuming the maximum allowable hydrogen-bond interaction. In the table below "Nuc. Stem" is the number of nucleotides within the stem region of the hairpin and does notinclude the nucleotides involved in end-fraying or loop formation; "Nuc. Other" are the number of nucleotides comprising the loop and end-fraying; "H-bonds" was calculated with the assumption of 3 hydrogen-bonds per GC basepair, and 2 hydrogen-bonds per G.G or T.T mismatch, one hydrogen bond per A.A or C.C mismatch; ΔH units are in kcal/mol, and describe on a theoretical "back of envelope" basis the maximum hydrogen bond energy; Enthalpy/nucleotide calculated in terms of the total strand length in nucleotides.

A

	ΔH/nuc	ΔH/nuc	ΔH/nuc	ΔH/nuc
Repeat	**CAG**	**CCG**	**CGG**	**CTG**
4	2.1	no fit	1.6	no fit
10	0.86	1.3	no fit	1.76
15	0.66	0.66	1.4	1.04
25	0.54	0.36	0.37	0.55

B

Theoretical, (CGG)n and (CTG)n					
Length	**Nuc.Stem**	**Nuc.Other**	**H-bonds**	**ΔH**	**ΔH/nuc.**
12	4	8	6	18	1.5
30	22	8	30	90	3
45	38	7	51	153	3.4
75	68	7	91	273	3.6
Theoretical, (CAG)n and (CCG)n					
Length	**Nuc.Stem**	**Nuc.Other**	**H-bonds**	**ΔH**	**ΔH/nuc.**
12	4	8	6	18	1.5
30	22	8	27	81	2.7
45	38	7	45	135	3
75	68	7	80	240	3.2

is lower for the longer sequences than for the shorter sequences. If the total enthalpy were the sum if the individual enthalpies for each triplet repeat unit, and assuming the structure assumes a conformation that allows the maximum amount of hydrogen–bonding and base stacking interactions, then the enthalpy per nucleotide should remain constant as function of length. This suggests long sequences are inherently biased towards instability. The data trend predicts that eventually the stability of these sequences will reach a plateau, beyond which no net gain in stability can be achieved.

4.3.2 Thermodynamic Parameters by Calorimetry

4.3.2.1 Comparison of Oligomer and Duplex Thermodynamic Parameters (Oligomer Concentration Adjusted)

The thermodynamic parameters for the (CNG) triplet repeat sequences, where N = A, C, G, or T and n = 4, 5, 10, 15, and 25, in conjunction with the complementary triplet repeat duplexes, (CNG)n/(CNG)n where N = A/T or C/G and n = 4, 10, 15, or 25 were determined with differential scanning calorimetry in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH7.0, 100 mM NaCl, with DNA concentrations between 20 μM and 100 μM. The excess heat capacity versus temperature data were collected and processed as previously described to obtain the transition melting temperature, the molar enthalpy, entropy, and Gibbs free energy for both the "structure" to single-strand transition in terms of the oligomer concentration. The calorimetrically derived thermodynamic parameters were considerably more accurate than the derived van't Hoff analysis parameters determined by the so-called "UV-melt" measurements for several reasons. First, the thermodynamic parameters derived by the calorimetry were model-independent (no

assumptions made about the state of the system) and independent of the requirement that only changes in absorbance reflected changes in the denaturation state. For the CCG4 and CTG4 oligomers, the van't Hoff thermodynamic parameters could not be derived due to the shallow nature of the transition. With DSC, the enthalpies were obtained. Third, the problems associated with van't Hoff analysis of non-ideal data were eliminated. For example, minor variations in the linear range of the van't Hoff plot could have vastly different results on the transition enthalpy (data not shown). Thus, where to place the linear range, over a variety of options, became a crucial decision in data processing. In addition, the thermograms contained less noise or measurement artifacts, such as biphasic melt profiles that were common to the transitions for the duplex melts and 75-mers. This eliminated the ambiguities associated with the van't Hoff analysis of such "UV-melt" curves. Finally, the observed enthalpy trends follow the observed structural changes by circular dichroism spectroscopy.

Due to idiosyncrasies of the calorimetric software ("CpCalc" by Calorimetric Science Corporation), the reported thermodynamic parameters were expressed as the enthalpy, entropy and free energy for the hairpin (HP) to single-strand (SS) transition. The implied direction meant that, assuming the SS-state was isoenergetic across all oligomers of a series, the more favorable enthalpy was achieved with larger (more positive) ΔH values, and that favorable entropy was achieved with smaller (less positive) ΔS values. Similarly, the free energy or stability relationships were reversed in a similar manner. The stability of the oligomers became favorable with increased positive magnitudes. This direction is opposite to that reported for the UV-thermal melts, where a more

negative enthalpy term indicated increased interactions, and a more positive entropy term indicated increased disorder.

The (CAG)n oligomers, where n = 4, 5, 10, 15, or 25 were studied by differential scanning calorimetry within the concentration range of 200 μM, for the 12-mer, to 50 μM, for the 75-mer, and produced the following trends as a function of length (Table 26, Figure 69). First, the enthalpy term became more favorable, from 5.5 kcal/mol with CAG4, to 114 kcal/mol with CAG25. This observation parallels the CD spectroscopic data, where the height of the peak at 275 nm increased in height as function of oligomer length. This trend is consistent with the UV-melt data that indicated that enthalpy became incrementally more favorable as function of oligomer length. This trend is not consistent with the CD spectra adjusted for nucleotide concentration, which suggests an overall decrease in the base nucleotide interactions. In light of these observations, the data suggests that while the base stacking interactions per nucleotide decrease, the amount of hydrogen bonding interactions increase. Over this same range the entropy became less favorable, increasing from 16 cal/(mol K), for the 12-mer, to 340 cal/(mol K) for the 75-mer. The free energy for the HP to SS transitional direction became more favorable, from 0.4 kcal/mol, for the 12-mer, to 9.0 kcal/mol for the 75-mer. Overall, the thermodynamic trends indicated that the stability of the (CAG)n oligomers increased as a function of length, and that the increases were enthalpically driven. Across the entire range of data, enthalpy/entropy compensation was observed (Figure 70). As the enthalpy became favorable, the entropy became unfavorable with the net effect that the stability of the oligomer was driven by the enthalpic contribution. The melting point for all

Table 26: The DSC-derived thermodynamic parameters for the (CNG)n oligomers adjusted for [oligomer].

The thermodynamic parameters of the (CNG)n oligomers, derived from differential scanning calorimetry measurements, are shown below as a function of duplex length. For all thermodynamic parameters listed below the assumed direction is the hairpin to single-strand transition. All parameters below are in terms of oligomer or duplex concentration.

Oligomer	Length	ΔH	ΔG_{37C}	ΔS	T_m
	nucleotides	kcal/mol	kcal/mol	cal/(mol K)	°C
CAG4	12	5.5	0.4	16.4	61.3
CAG5	15	16.9	1.5	49.8	65.8
CAG10	30	44.7	3.4	132.0	62.5
CAG15	45	56.4	4.3	167.8	62.1
CAG25	75	114.0	9.0	339.0	62.2
Oligomer	**Length**	**ΔH**	**ΔG_{37C}**	**ΔS**	**T_m**
	nucleotides	kcal/mol	kcal/mol	cal/(mol K)	°C
CCG4	12	5.7	0.2	17.8	52.9
CCG5	15	8.7	0.5	26.5	57.0
CCG10	30	57.0	2.4	177.0	50.0
CCG15	45	76.6	3.5	235.8	51.8
CCG25	75	101.8	4.0	315.3	48.8
Oligomer	**Length**	**ΔH**	**ΔG_{37C}**	**ΔS**	**T_m**
	nucleotides	kcal/mol	kcal/mol	cal/(mol K)	°C
CGG4	12	23.2	1.4	70.3	57.7
CGG5	15	9.5	0.7	28.5	58.3
CGG10	30	92.8	11.9	261.0	82.0
CGG15	45	147.8	19.6	413.3	84.8
CGG25	75	148.5	20.5	411.3	87.8
Oligomer	**Length**	**ΔH**	**ΔG_{37C}**	**ΔS**	**T_m**
	nucleotides	kcal/mol	kcal/mol	cal/(mol K)	°C
CTG4	12	1.9	0.0	6.0	53.2
CTG5	15	23.4	0.6	73.8	44.3
CTG10	30	77.1	6.4	228.0	64.8
CTG15	45	126.0	10.9	371.3	65.9
CTG25	75	165.0	14.5	486.0	66.7

Figure 69: Representative DSC thermograms of the (CAG)n oligomers.

Representative DSC thermograms for the (CAG)n series of oligomers, where n = 4, 10, 15, 0r 25, is shown below. **A.** The CAG4 oligomer. **B.** The CAG10 oligomer. **C.** The CAG15 oligomer. **D.** The CAG25 oligomer.

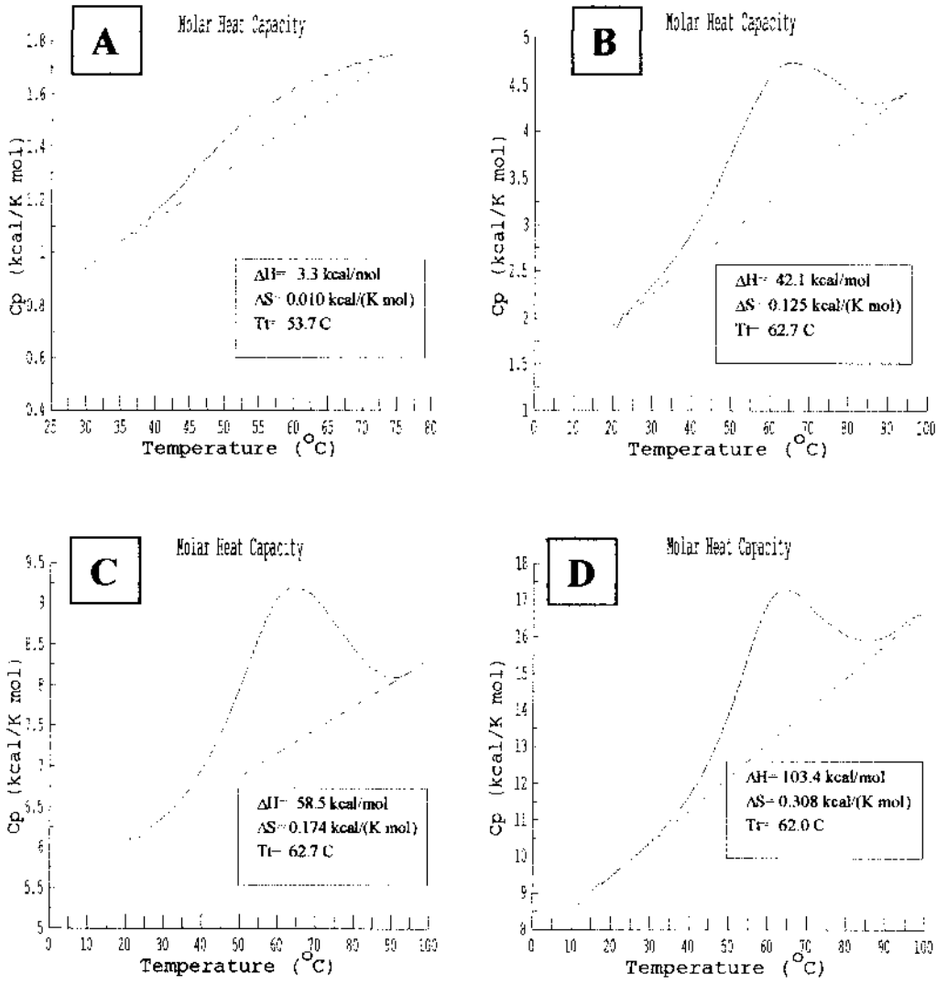
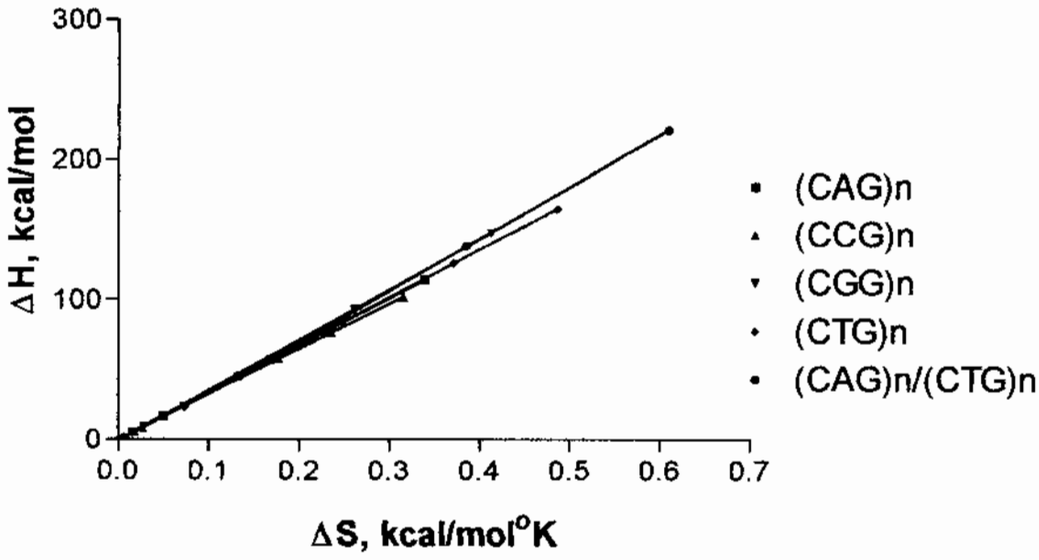


Figure 70: Enthalpy/Entropy plot of the (CNG)n oligomers, and (CAG)n/(CTG)n duplex.

The enthalpy for each oligomer is plotted as function of entropy for the hairpin (or other structure) to single strand transition - favorable entropy is associated with smaller values of ΔS, and favorable enthalpy with large values of ΔH. The inset below shows the linear regression parameters for each sample.



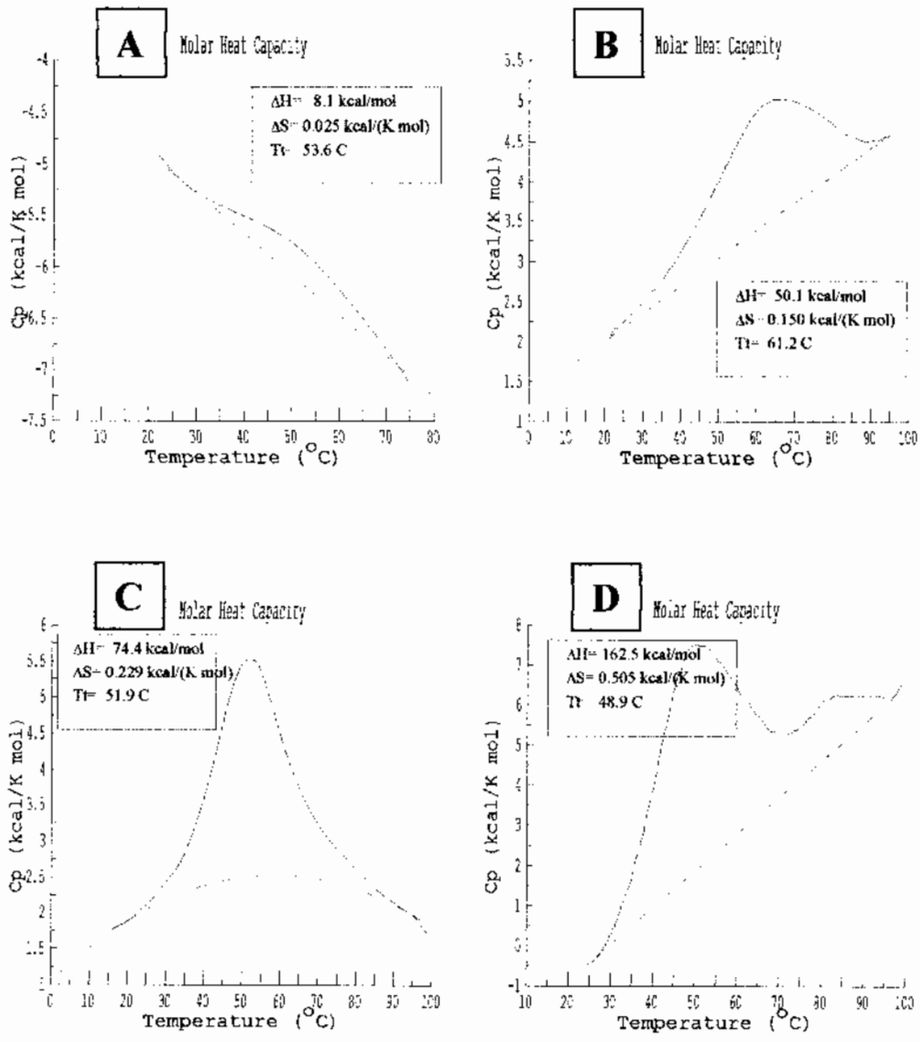
Oligomer	m	b	r2
(CAG)n	336	0.184	1.00
CCG)n	323	0.055	1.00
(CGG)n	364	-1.99	1.00
(CTG)n	342	-0.805	1.00
(CAG)n/(CTG)n	371	-5.059	1.00

oligomers studied appeared to reach a plateau, which is in contrast to the observed stability trend for the CAG oligomers.

The (CCG)n oligomers, where n = 4, 5, 10, 15, or 25, were also measured within the same concentration range, with the same techniques as used for the CAG oligomers (Table 26, Figure 71). The enthalpy becomes more favorable, from 5.7 kcal/mol for the 12-mer to 102 kcal/mol for the 75-mer. The observed trend in enthalpy is consistent with the CD spectral data that indicated an increase in base stacking interactions, by the increase the height of the peak near 285 nm, as a function of length. The observed calorimetric enthalpy trends are in contrast to the trends obtained by van't Hoff analysis of the UV-melt data. Overall the enthalpy became more favorable, but the magnitude of the change was small and possibly within the experimental error of the assay. The calorimetric results leave no doubt as to the direction and magnitude of the enthalpic changes as a function of length. The entropy for the CCG oligomers becomes less favorable, from 18 cal/(mol K) for the 12-mer, to 315 cal/(mol K) for the 75-mer. As observed for the CAG oligomers, enthalpy/entropy compensation occurred throughout the entire range of lengths studied (Figure 70). The magnitude of the entropy however, was not great enough to overcome the enthalpic contribution. The stability mirrored the trends observed for the enthalpy. The free energy began at 0.2 kcal/mol for the 12-mer, and became more favorable to 4.0 kcal/mol for the 75-mer. Overall, the stability of the CCG oligomers was enthalpically driven.

Figure 71: Representative DSC thermograms of the (CCG)n oligomers.

Representative DSC thermograms for the (CCG)n series of oligomers, where n = 4, 10, 15, 0r 25, is shown below. **A.** The CCG4 oligomer. **B.** The CCG10 oligomer. **C.** The CCG15 oligomer. **D.** The CCG25 oligomer.

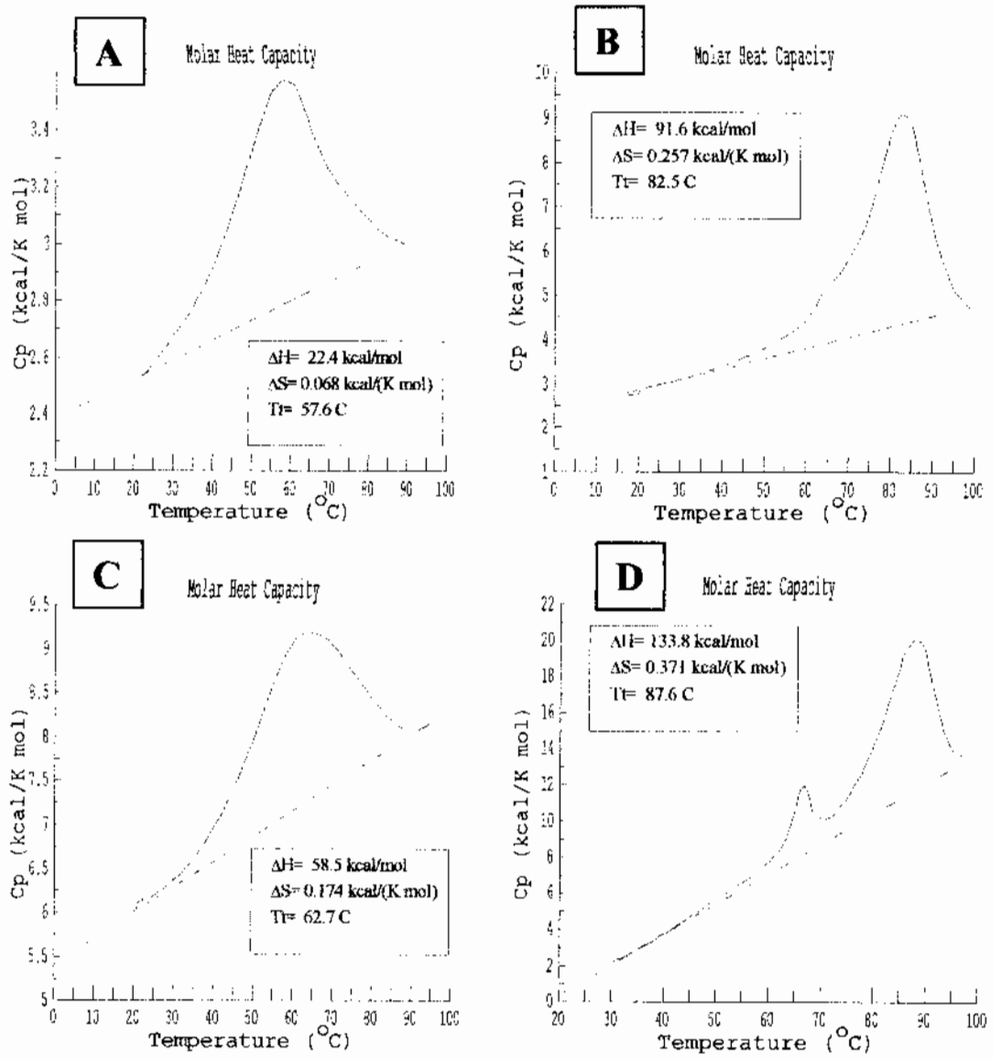


The (CGG)n oligomers, where n = 4, 5, 10, 15, or 25, were studied by differential scanning calorimetry within the concentration range of 150 μM, for the 12-mer, to 75 μM, for the 75-mer, and produced the following trends as a function of length (Table 26, Figure 72). The enthalpy began from 23 kcal/mol for the 12-mer and reached a plateau of approximately 150 kcal/mol for the 45-mer and 75-mer of the (CGG)n oligomers. This trend is inconsistent with the trend observed with the UV-melt data that indicated a slight peak at approximately 45 nucleotides (Table 23). This data however is consistent with the CD spectral data adjusted on the basis of oligomer concentration. The peak at 282 nm and the trough at 250 nm incrementally change as a function of length. The peak at 240 nm however, increased up to the 45-mer but then appeared to plateau beyond 45 nucleotides, suggesting a possible structural link for the thermodynamic data (Figure 41). The entropy for these oligomers became less favorable, from 70 cal/(mol K) for the 12-mer to 411 cal/(mol K) for the 75-mer. There was enthalpy/entropy compensation occurring, where the entropy term became more favorable with increasing unfavorability of the enthalpy (Figure 70). Overall, the change in entropy was not sufficient to drive the directionality of the stability. The free energy of the stability, on an oligomer basis, reflected the changes in the enthalpy and began from 1.4 kcal/mol for the 12-mer to 20 kcal/mol for the 75-mer. The observed trends were inconsistent with the UV-melt data, and but consistent with the CD spectral data at 240 nm adjusted to oligomer concentration (Figure 41).

The (CTG)n oligomers, where n = 4, 5, 10, 15, or 25, were studied within the same concentration range as all other samples and produced the following trends as a function

Figure 72: Representative DSC thermograms of the (CGG)n oligomers.

Representative DSC thermograms for the (CGG)n series of oligomers, where n = 4, 10, 15, 0r 25, is shown below. **A.** The CGG4 oligomer. **B.** The CGG10 oligomer. **C.** The CGG15 oligomer. **D.** The CGG25 oligomer.

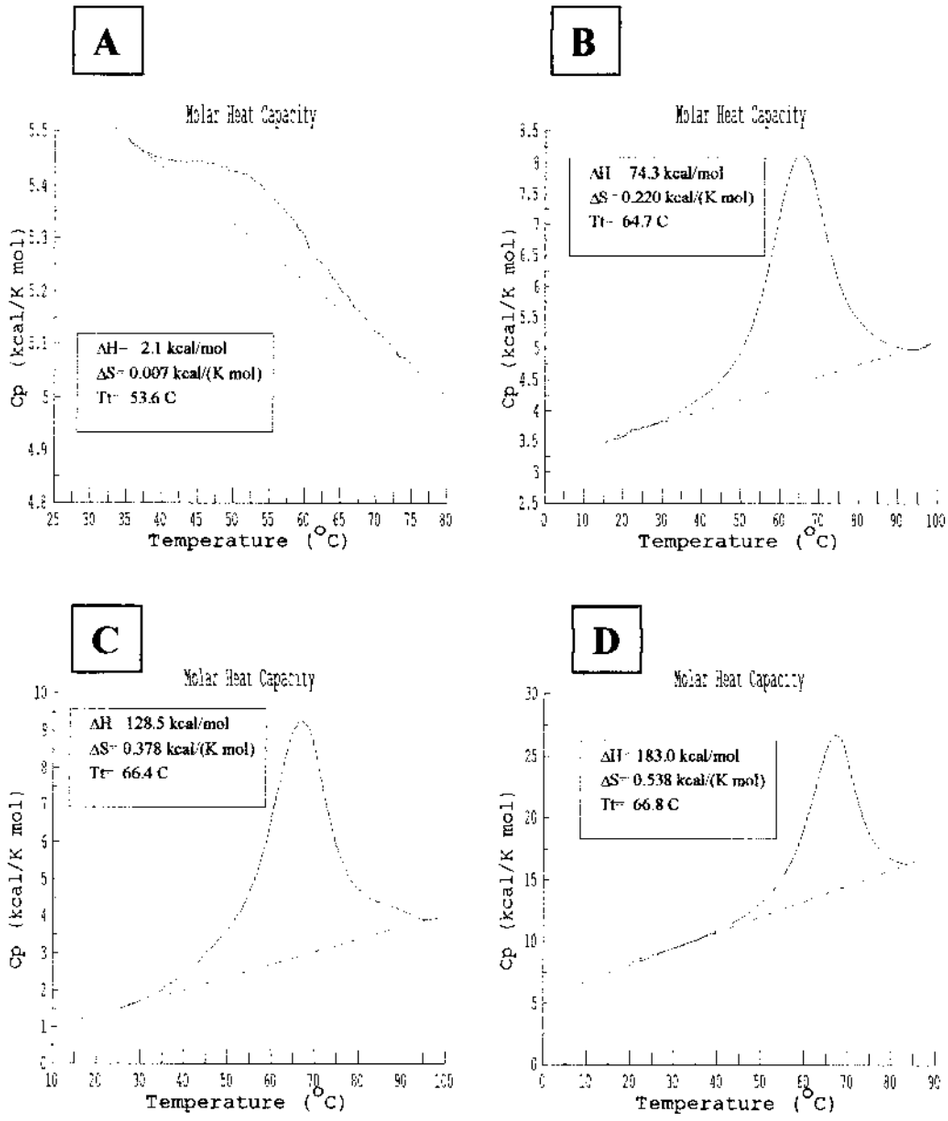


of length (Table 26, Figure 73). The enthalpy incrementally increased, becoming more favorable, from 2 kcal/mol for the 12-mer to 165 kcal/mol for the 75-mer. This result is in disagreement with the enthalpy trend observed for the UV-melt data set, which indicated a net loss of enthalpy with increasing length. The increase in enthalpy is consistent with the CD spectral data, adjusted for oligomer concentration, which indicated the extent of interactions should increase with increasing oligomer length (Figure 40). The entropy for these oligomers gave the expected trend and was consistent with all other oligomers in this study. The entropy became less favorable with increasing favorability of the enthalpy, and went from 6.0 cal/(mol K) for the 12-mer , to 486 cal/(mol K) for the 75-mer. The enthalpy/entropy compensation phenomenon (Figure 70) was also observed for these sequences. The net result of the compensation was, as observed for all other oligomers in this study, the entropy was significant enough to drive the directionality of free energy. The free energy followed the same general trend as the enthalpy, and became more favorable, from 0.1 cal/(mol K) for the 12-mer to 15 cal/(mol K) for the 75-mer. Like the enthalpy, the observed trend of the free energy trend is not consistent with the UV-melt data that suggested the free energy became less favorable as a function of length (Table 23).

The equi-length complementary duplexes of the (CNG)n/(CNG)n oligomers, where N = A/T or C/G and n = 4, 10, 15, or 25, were studied by differential scanning calorimetry. The melting transitions for the CCG/CGG duplex series were too high for accurate measurement by DSC, and were therefore not integrated, as in the case of CCG10/CGG10 or not measured as in the case of CCG15/CGG15, and CCG25/CGG25

Figure 73: Representative DSC thermograms of the (CTG)n oligomers.

Representative DSC thermograms for the (CTG)n series of oligomers, where n = 4, 10, 15, 0r 25, is shown below. **A.** The CTG4 oligomer. **B.** The CTG10 oligomer. **C.** The CTG15 oligomer. **D.** The CTG25 oligomer.



(Figure 74). The remaining duplex set, the (CAG)n/(CTG)n duplexes, were within the realm of accurate measurement and were promptly quantified by DSC. The heat capacity of the (CAG)n/(CTG)n duplexes, where n = 4, 10, 15, or 25, were measured by DSC and these measurements produced the following thermodynamic trends (Figure 75). The thermal transition for the CAG25/CTG25 oligomer was too high for accurate quantitation, and was not entered into the data set for analysis. The enthalpy for the CAG4/CTG4 through CAG15/CTG15 samples became more favorable, and increased from 50 kcal/mol for the 12-bp duplex, to 220 kcal/mol for the 45-bp duplex. The observed enthalpy trend is not consistent with the same parameter measured by UV-melt techniques (Table 24). The enthalpy trend for the UV-melt data indicated that the enthalpy became less favorable, from –70 kcal/mol to –30 kcal/mol, as a function of length. The calorimetric data however is consistent with the CD spectral data that was adjusted for oligomer concentration (Figure 42). The CD data suggested that the overall base stacking interactions, as indicated by the increased peak height at 286 nm and the increased depth of the trough at 259 nm, increased with increasing oligomer length. The calorimetrically derived entropy for these duplexes became less favorable with increasing duplex length, beginning from 155 cal/(mol K) for the 12-bp duplex to 610 cal/(mol K) for the 45-bp duplex. While entropy/enthalpy compensation occurred (Figure 70), the observed stability was driven by the favorable enthalpy. The stability trend, like the observed enthalpy trend, became more favorable from 4 kcal/mol for the 12-bp duplex to 31 kcal/mol for the 45-bp duplex. The observed free energy trend is not in agreement with the free energy trend determined by van't Hoff analysis of the UV-melt curves. The free energy trend by UV-melt techniques suggested a peak around 45 nucleotides, with a

Figure 74: Representative DSC thermograms of the (CCG)n/(CGG)n duplexes.

Representative DSC thermograms for the (CAG)n/(CTG)n series of duplexes, where n = 4, 10, 15, 0r 25, is shown below. **A.** The CAG4/CTG4 duplex. **B.** The CAG10/CTG10 duplex. **C.** The CAG15/CTG15 duplex.

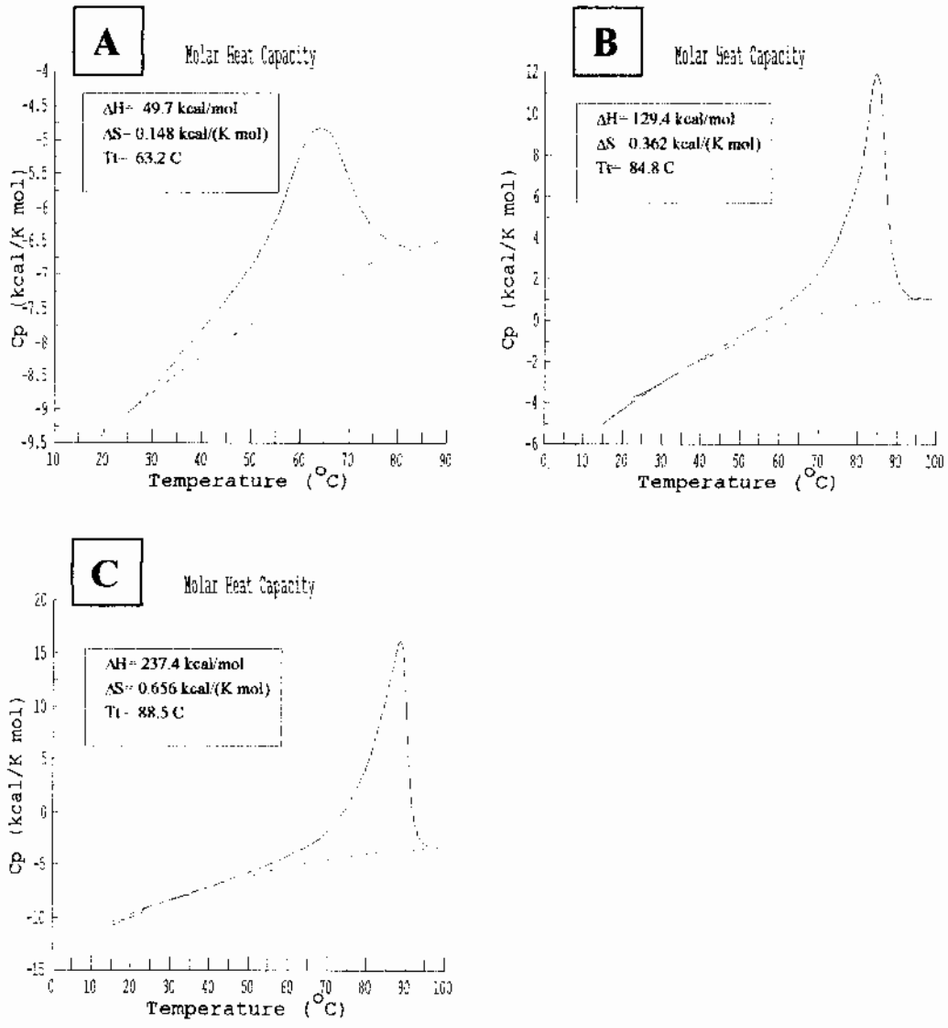


Figure 75: Representative DSC thermograms of the (CAG)n/(CTG)n duplexes.

Representative DSC thermogram for the (CCG)4/(CGG)4 duplex is shown below.

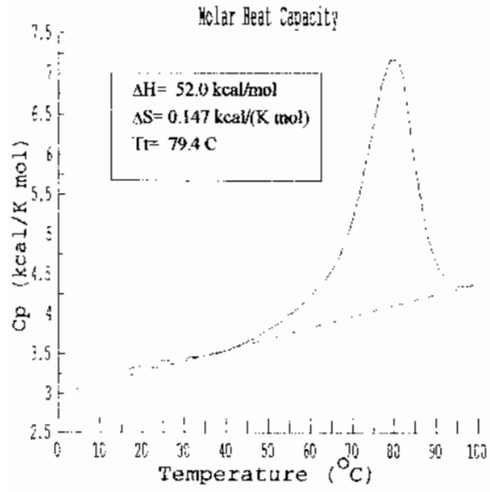


Table 27: The DSC-derived thermodynamic parameters for the (CNG)n/(CNG)n duplexes adjusted for [duplex].

The thermodynamic parameters of the (CNG)n/(CNG)n oligomers, derived from differential scanning calorimetry measurements, are shown below as a function of duplex length and were adjusted for duplex concentration. For all thermodynamic parameters listed below the assumed direction is the hairpin to single-strand transition. Those samples with incomplete transitions, where the endpoint of the thermogram could not be clearly defined, are labeled "inc. trans." Those samples with an apparent melting transition that was near or above 95 °C are denoted with "$T_m > 95C$".

Duplex	Length	ΔH	ΔG_{37C}	ΔS	T_m
	base-pairs	kcal/mol	kcal/mol	cal/(mol K)	°C
CAG4/CTG4	12	52.2	4.0	155.3	63.4
CAG10/CTG10	30	138.1	18.5	385.8	84.8
CAG15/CTG15	45	220.4	31.5	609.5	88.4
CAG25/CTG25	75	T_m >95 C	T_m >95 C	T_m >95 C	T_m >95 C
Duplex	**Length**	**ΔH**	**ΔG_{37C}**	**ΔS**	**T_m**
	base-pairs	kcal/mol	kcal/mol	cal/(mol K)	°C
CCG4/CGG4	12	51.2	6.2	145.3	79.3
CCG5/CGG5	15	inc. trans	inc. trans	inc. trans	96.0
CCG10/CGG10	30	T_m >95 C	T_m >95 C	T_m >95 C	T_m >95 C
CCG15/CGG15	45	T_m >95 C	T_m >95 C	T_m >95 C	T_m >95 C
CCG25/CGG25	75	T_m >95 C	T_m >95 C	T_m >95 C	T_m >95 C

subsequent decrease in stability for the 75-bp duplex. It is possible that the observed trend by van't Hoff analysis of the UV-melt curves is an artifact of the data quality and data reduction process.

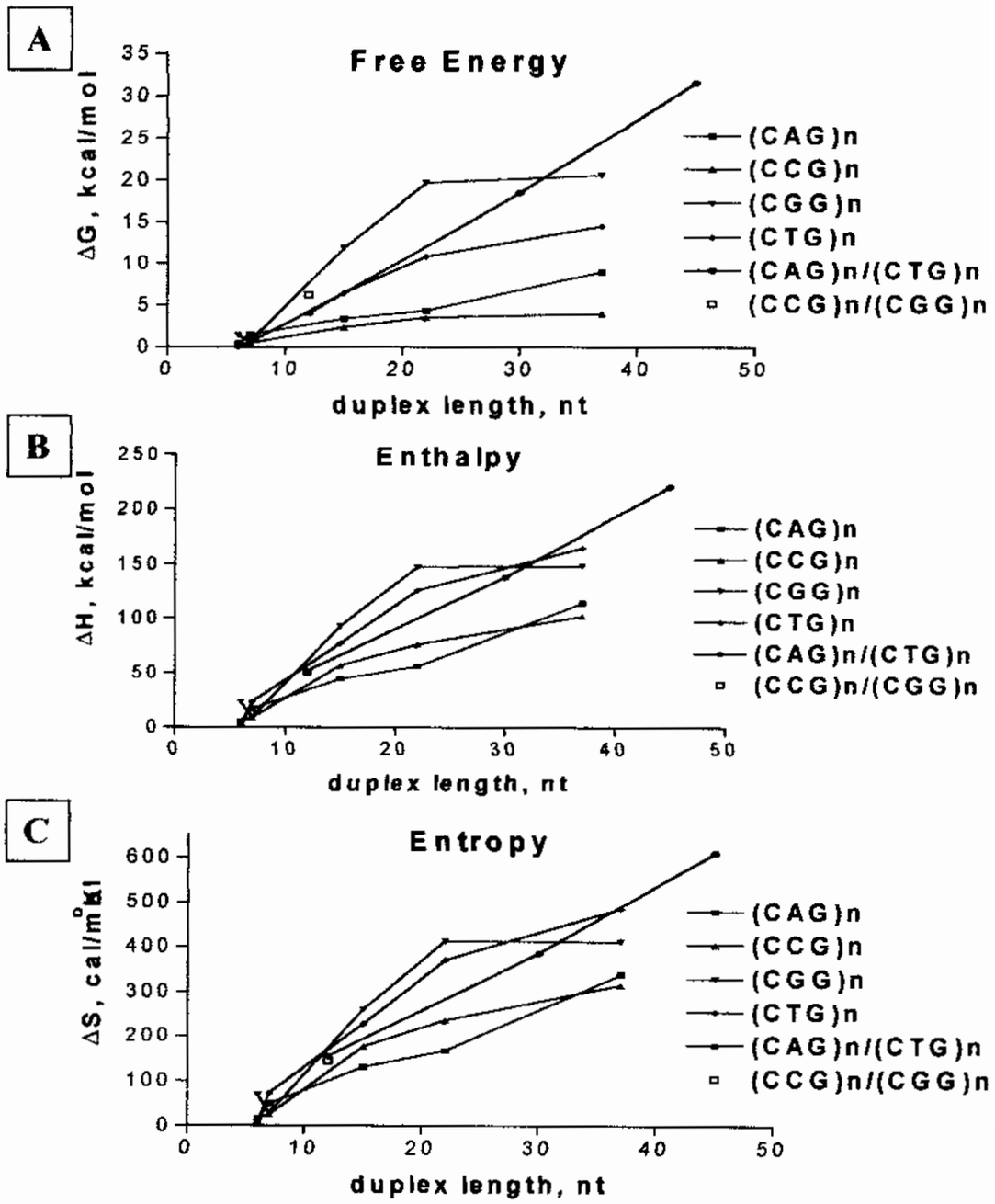
A plot of the oligomer and duplex stability, enthalpy and entropy provided an interesting and meaningful juxtaposition of the two data sets that revealed trends not apparent by inspection of tabular thermodynamic data (Figure 76). Comparison of the thermodynamic parameters between the solitary oligomers (CAG)n and (CTG)n and the approximated corresponding duplex (CAG)n/(CTG)n showed that the stability of the duplex is equal to or greater than the oligomer (Table 26, Table 27, Figure 76). For example, if the 30-mers of either CAG30 or CTG30 were to form a blunt-ended hairpin, the stem of the hairpin would contain approximately 15-basepairs in a B-like DNA conformation. The corresponding duplex CAG15/CTG15 had more stability in terms of free energy than the corresponding hairpin as determined by a plot of the free energy vs. duplex length (Figure 76). However, the enthalpy for the same sequences showed that the (CTG)n oligomers in the hairpin conformation had slightly more base stacking, and/or hydrogen-bond interactions than the corresponding duplex at shorter lengths. Eventually, however, the enthalpy of the duplex became larger, due to a plateau in the (CTG)n enthalpy, which also translated into a plateau in the (CTG)n free energy as a function of length. The (CAG)n oligomers one the other hand, displayed no such tendencies for greater enthalpy or free energy than the corresponding duplex.

subsequent decrease in stability for the 75-bp duplex. It is possible that the observed trend by van't Hoff analysis of the UV-melt curves is an artifact of the data quality and data reduction process.

A plot of the oligomer and duplex stability, enthalpy and entropy provided an interesting and meaningful juxtaposition of the two data sets that revealed trends not apparent by inspection of tabular thermodynamic data (Figure 76). Comparison of the thermodynamic parameters between the solitary oligomers (CAG)n and (CTG)n and the approximated corresponding duplex (CAG)n/(CTG)n showed that the stability of the duplex is equal to or greater than the oligomer (Table 26, Table 27, Figure 76). For example, if the 30-mers of either CAG30 or CTG30 were to form a blunt-ended hairpin, the stem of the hairpin would contain approximately 15-basepairs in a B-like DNA conformation. The corresponding duplex CAG15/CTG15 had more stability in terms of free energy than the corresponding hairpin as determined by a plot of the free energy vs. duplex length (Figure 76). However, the enthalpy for the same sequences showed that the (CTG)n oligomers in the hairpin conformation had slightly more base stacking, and/or hydrogen-bond interactions than the corresponding duplex at shorter lengths. Eventually, however, the enthalpy of the duplex became larger, due to a plateau in the (CTG)n enthalpy, which also translated into a plateau in the (CTG)n free energy as a function of length. The (CAG)n oligomers one the other hand, displayed no such tendencies for greater enthalpy or free energy than the corresponding duplex.

Figure 76: Comparison of the oligomer stability and duplex stability.

The oligomer-concentration adjusted free energy (**A**), enthalpy (**B**), and entropy (**C**) of the (CNG)n oligomers and the (CNG)n/(CNG)n duplexes. For all oligomers shown the "duplex" length of the hairpin is the length of the stem approximated by the total nucleotide length divided by two.



The overlay of the duplex and oligomer stabilities as a function of length also revealed a new trend that was not apparent from the tabular data. It is also interesting to note that for all solitary oligomers capable of forming hairpin conformations, the stability, enthalpy and entropy displayed a tendency to plateau, whereas the corresponding complementary duplex displayed no such tendencies within the length region studied (Figure 76). While there were increases in the thermodynamic parameters as a function of length, the observed increases for the hairpin oligomers did not follow the same near-linear trajectory as observed for the corresponding duplex beyond 30 base-pairs. This suggests that while there are slight stability advantages to be gained with increasing length, the overall stability is not additive. Relative to the complementary duplex, the longer the oligomer, the less stability of the oligomer resembles that of the duplex. This observation is consistent with a strand that becomes comparatively less stable, in a thermodynamic sense, as the length of the triplet repeat region increases thus increasing the probability for alternative structure formation at longer lengths. With respect to triplet repeat sequence DNA, this is the first documented case where there is a connection between the length and the underlying disease mechanism.

Comparison of overall oligomer to oligomer stability shows the following stability order with respect to the (CNG)15 stabilities (Figure 76, Table 26). The observed stability order, in terms of the free energy of the HP to SS transition is, most stable to least stable, CGG > CTG > CAG > CCG. This is in contrast to the enthalpy trend that shows a reversal in interactions for the CCG and CAG oligomers, form most to least stable, CGG > CTG > CCG > CAG. The observed reversal is due most likely to the relatively large

contribution made by the entropic term of the (CCG)n oligomers that serves to reduce the overall stability.

The observed slopes of the enthalpy versus entropy plot (Table 28, Figure 70) were within the same order of magnitude, and followed the same trend observed for the thermodynamic stability in terms of the free energy. As shown previously, (above paragraph) the stability trend from most to least stable was CGG > CTG > CAG > CCG. The trend observed for the slopes of the enthalpy versus entropy plot proceeded, from highest slope to lowest slope, CGG > CTG > CAG > CCG. Considering the similarity between the slopes the data were further analyzed by Student's t-test. Statistical analysis of the enthalpy and entropy plots with a 0.01 confidence limit showed that in all cases the

Table 28: Slopes of the enthalpy versus entropy plot .

The slopes of the enthalpy versus entropy plots show that there is a small but statistically significant difference between the sequence classes of strands. The most stable solitary oligomer, (CGG)n has more enthalpy per nucleotide than the least stable oligomer (CCG)n. In each t-test comparison, the slope of the line under examination was compared to the slope of the line immediately below. For example, the slope of the CAG/CTG line was compared to the slope of the (CGG)n line; the slope of the (CGG)n line was compared to the slope of the (CTG)n line until the last comparison was drawn between the (CAG)n and (CCG)n line slopes. In each plot the mean slope is an average of four determinations. The t-value for a statistically significant event given 6 degrees of freedom and a confidence limit of 0.01 is 3.7.

Oligomer or Duplex	ΔH / ΔS slope, mean (J/mol)/(J/molK)	Standard deviation	t
CAG/CTG	371	+/- 3	7
(CGG)n	360	+/- 0.7	51
(CTG)n	342	+/- 0.1	14
(CAG)n	336	+/- 0.8	22
(CCG)n	323	+/- 0.8	-

lines were statistically different from each other. This result demonstrates a clear

relationship between the enthalpy and the observed stability of the oligomers. The (CGG)n series of oligomers are more stable than the (CCG)n oligomers due to increased enthalpic contributions.

4.3.2.2 Comparison of Oligomer and Duplex Thermodynamic Parameters (Nucleotide-Concentration Adjusted)

The thermodynamic parameters for the (CNG) triplet repeat sequences, where N = A, C, G, or T and n = 4, 5, 10, 15, and 25, in conjunction with the complementary triplet repeat duplexes, (CNG)n/(CNG)n where N = A/T or C/G and n = 4, 10, 15, or 25 were determined with differential scanning calorimetry in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH7.0, 100 mM NaCl, with DNA concentrations between 20 μM and 100 μM. The excess heat capacity versus temperature data were collected and processed as previously described to obtain the transition melting temperature, the molar enthalpy, entropy, and Gibbs free energy for the "structure" to single-strand transition in terms of the nucleotide concentration. Comparison of the enthalpies adjusted for nucleotide concentration gives an indirect measure of the thermodynamic changes occurring at the nucleotide level that is averaged across the entire oligomer (Table 29). Comparison of the enthalpic trends to the CD spectral data provided a connection between the thermodynamic observable, the enthalpy, and the structural observable, the change in base stacking interactions. Such comparisons revealed thermodynamic trends that were, in general, similar to those observed with the CD spectra adjusted for nucleotide concentration, and suggested that the observed changes in stability, were due to changes in the base stacking interactions.

The nucleotide adjusted enthalpy of the (CAG)n oligomers, where n = 4, 10, 15, or 25, produced a trend with some fluctuation between the 30-mer and the 75-mer that made conclusive data reduction difficult (Table 29). The enthalpy per nucleotide increased from approximately 0.5 kcal/mol_{nuc} for CAG4 to 1.5 kcal/mol_{nuc} for CAG10. The enthalpy per nucleotide then decreased slightly to 1.24 kcal/mol_{nuc} for CAG15 but then increased to approximately 1.5 kcal/mol_{nuc} for CAG25. The calorimetric data suggests that either there is a slight decrease in the enthalpy/nucleotide (if the CAG25 data point is errant) or the enthalpy/nucleotide reaches a plateau (if the CAG15 data point is errant). The CD spectral data suggests that the base stacking interactions for the (CAG)n oligomers reach a maximum at approximately 45 nucleotides (Figure 51, A). In

Table 29: The DSC-derived thermodynamic parameters for the (CNG)n oligomers adjusted for [nucleotide]

The thermodynamic parameters of the (CNG)n oligomers, derived from differential scanning calorimetry measurements, are shown below as a function of duplex length. For all thermodynamic parameters listed below the assumed direction is the hairpin to single-strand transition. All thermodynamic parameters below are in terms of the nucleotide concentration

Oligomer	Length	ΔG_{37C}	ΔH	ΔS
	nucleotides	kcal/mol$_{nuc.}$	kcal/mol$_{nuc.}$	cal/mol$_{nuc.}$·°K
CAG4	12	0.03	0.46	1.36
CAG5	15	0.10	1.13	3.32
CAG10	30	0.11	1.49	4.40
CAG15	45	0.10	1.25	3.73
CAG25	75	0.12	1.52	4.52
Oligomer	**Length**	ΔG_{37C}	ΔH	ΔS
	nucleotides	kcal/mol$_{nuc.}$	kcal/mol$_{nuc.}$	cal/mol$_{nuc.}$·°K
CCG4	12	0.02	0.48	1.48
CCG5	15	0.03	0.58	1.77
CCG10	30	0.08	1.90	5.90
CCG15	45	0.08	1.70	5.24
CCG25	75	0.05	1.36	4.20
Oligomer	**Length**	ΔG_{37C}	ΔH	ΔS
	nucleotides	kcal/mol$_{nuc.}$	kcal/mol$_{nuc.}$	cal/mol$_{nuc.}$·°K
CGG4	12	0.12	1.93	5.85
CGG5	15	0.05	0.64	1.90
CGG10	30	0.40	3.09	8.70
CGG15	45	0.44	3.28	9.18
CGG25	75	0.27	1.98	5.48
Oligomer	**Length**	ΔG_{37C}	ΔH	ΔS
	nucleotides	kcal/mol$_{nuc.}$	kcal/mol$_{nuc.}$	cal/mol$_{nuc.}$·°K
CTG4	12	0.00	0.16	0.50
CTG5	15	0.04	1.56	4.92
CTG10	30	0.21	2.57	7.60
CTG15	45	0.24	2.80	8.25
CTG25	75	0.19	2.20	6.48

combination with the CD spectral data the calorimetric data indicates that the enthalpy/nucleotide increases due to a slight increase in the base stacking interactions and/or hydrogen-bond formation and a decrease in base stacking interactions as the length of the oligomer increases.

The nucleotide concentration adjusted enthalpy of the (CCG)n oligomers began at 0.48 kcal/mol_{nuc} for the 12-mer and reached a maximum at approximately 30 nucleotides with 1.9 kcal/mol_{nuc} (Table 29). The enthalpy then decreased to 1.36 kcal/mol with the 25-mer. These trends paralleled those observed for the CD spectral data, which had shown an increase in the base stacking interactions per nucleotide as the length increased up to approximately the 45-mer (Figure 51, B). The base stacking interactions, adjusted on a nucleotide basis, then decreased for the 75-mer. The data suggests that the observed decrease in enthalpy/nucleotide may be related to a decrease in the base stacking interactions that help hold the (CCG)n oligomers together.

The thermodynamic trends for the (CGG)n oligomers also exhibit behavior that is similar to that observed for the (CCG)n oligomers (Table 29). The enthalpy began from 1.9 kcal/mol_{nuc} for the 12-mer and then reached a maximum of 3.1-3.3 kcal/mol_{nuc} for the 30-mer and 45-mer respectively. The enthalpy/nucleotide then decreased to 2 kcal/mol_{nuc} for the 75-mer. Like the enthalpic data the CD spectral data indicated that the base stacking interactions, as indicated by the trough at approximately 250 nm and peak at 240 nm, reach a maximum around 30 nucleotides (Figure 50, A). The base stacking interactions then decrease, as shown by the decrease in the depth of the trough at 250 nm

and decrease in the height of the peak at 240 nm, once the oligomer achieves the length of 75 nucleotides. These data suggest that the loss of enthalpic stability per nucleotide may be related to the loss of base stacking interactions.

The enthalpy per nucleotide trends for the (CTG)n oligomers also follow the same trends previously observed for the (CCG)n and (CGG)n oligomers (Table 29). The enthalpy/nucleotide achieved the maximum value at approximately 45 nucleotides. The enthalpy began at 0.16 kcal/mol_{nuc} for the 12-mer, peaked at 2.8 kcal/mol_{nuc} for the 45-mer, and then decreased to 2.2 kcal/mol_{nuc} for the 75-mer. The CD spectral data also displayed similar and equally subtle trends. On a nucleotide concentration adjusted basis, the height of the peak at 286 nm reached a maximum at approximately 45 nucleotides (Figure 50, B). Like prior data, the data for the (CTG)n oligomers suggests that the loss of enthalpy on a nucleotide adjusted basis is due to a loss of base stacking interactions.

The (CAG)n/(CTG)n duplexes displayed trends that were consistent with the expected behavior for duplex DNA (Table 30). For duplex DNA, it was expected that the enthalpy per nucleotide would remain the same, or perhaps increase slightly, as function of length due to the addition of Watson-Crick hydrogen-bonded base-pairs to the duplex. Incremental additions of basepairs to a DNA duplex are expected in increase the enthalpy as a function of oligomer concentration, where the enthalpic contributions are additive, but maintain the enthalpy per nucleotide, where the enthalpic contributions are averaged across the entire oligomer. It possible however, with slight alterations in DNA structure that the enthalpic contribution per nucleotide could increase/decrease, slightly, as a

Table 30: The DSC-derived thermodynamic parameters for the (CAG)n/(CTG)n duplexes adjusted for [nucleotide].

The thermodynamic parameters of the (CAG)n/(CTG)n oligomers, derived from differential scanning calorimetry measurements, are shown below as a function of nucleotide concentration. For all thermodynamic parameters listed below the assumed direction is the hairpin to single-strand transition. Those samples with incomplete transitions, where the endpoint of the thermogram could not be clearly defined, are labeled "inc. trans."

Oligomer	Length	ΔG_{37C}	ΔH	ΔS
	base-pairs	kcal/mol$_{nuc.}$	kcal/mol$_{nuc.}$	cal/mol$_{nuc.}$ K
CAG4/CTG4	12	0.17	2.17	6.47
CAG10/CTG10	30	0.31	2.30	6.43
CAG15/CTG15	45	0.35	2.45	6.77
CAG25/CTG25	75	T_m >95 C	T_m >95 C	T_m >95 C

function of the length-induced structural changes. For the (CAG)n /(CTG)n duplexes the enthalpy per nucleotide increased, from 2.16 kcal/mol$_{nuc}$ for the 12-bp duplex to 2.44 kcal/mol$_{nuc}$ for the 45-bp duplex. The (CGG)n and (CTG)n oligomers displayed similar behaviors within the same range, but with a greater difference between the 12-mer and the 45-mer. The range of useable data did not extend to the 75-bp duplex because the melting transition of the CAG25/CTG25 duplex was too high to obtain useable data. Whether or not the (CAG)n/(CTG)n duplexes would have followed similar trends as the (CGG)n or (CTG)n oligomers remains undefined. The thermodynamic data suggests that the base stacking and hydrogen-bond interactions per nucleotide increase slightly, as a function of increasing base-pairs in the oligomer. In contrast, the CD spectral data, adjusted for nucleotide concentration, displayed trends that did not match the observed enthalpic tendency for increased stability (Figure 54). Both the height of the peak near 275 nm and the depth of the trough near 255 nm decreased in magnitude and suggested that the base stacking interactions per nucleotide decreased with length. This suggests,

possibly, the formation of a structure that, while increasing the overall hydrogen-bonding interactions decreased the base stacking interactions per nucleotide. The tendency for decreased interaction is mirrored with the CD spectra of these duplexes as a function of length – the base stacking interactions, which should show incremental increases as a function of length, appear to reach a plateau (Figure 42, A)

4.3.2.3 Comparison of Calorimetric and Modeled Calorimetric Data.

The calorimetrically derived excess heat capacity data were imported into the deconvolution modules of the "CpCalc" software, and the van't Hoff transition enthalpies were determined with a ""Two-State Scaled"" model. Previous trials with the "standard" two-state fit yielded results that were similar to the "Two-State Scaled" fits in terms of the magnitudes of the transitional enthalpies, but with marginally higher X^2 statistical values (data not shown). Fitting the data to a "DNA Melting Curve" model yielded enthalpies that were neither similar to the two-state fit nor had similar values for the chi-squared test (data not shown). In most examples, the fit line with the "DNA Melting Curve" was a poor description of the calorimetric data (data not shown). Therefore, data from the "Two-State Scaled" model were used for all subsequent modeling experiments.

For all oligomer data sets, the "Two-State Scaled" model with one transition was sufficient to fit the data for all 12-mers. With increasing length however, the complexity of the calorimetric peak shapes increased and became less gaussian in shape. The base of the peaks became broader with some transitional asymmetry due end-fraying or other phenomena in the pre-transition region (Figure 77, Figure 78, Figure 79, Figure 80). The

Figure 77: Representative van't Hoff modeling of the thermograms for the (CAG)n oligomers.

Representative "Two-State Scaled" data models for the (CAG)n series of oligomers, where n = 4, 10, 15, or 25 are shown below. For all data sets, the blue line represents the calorimetric data, and the red line represents the sum of the van't Hoff modeled data. **A.** The CAG4 oligomer. **B.** The CAG10 oligomer. **C.** The CAG15 oligomer. **D.** The CAG25 oligomer.

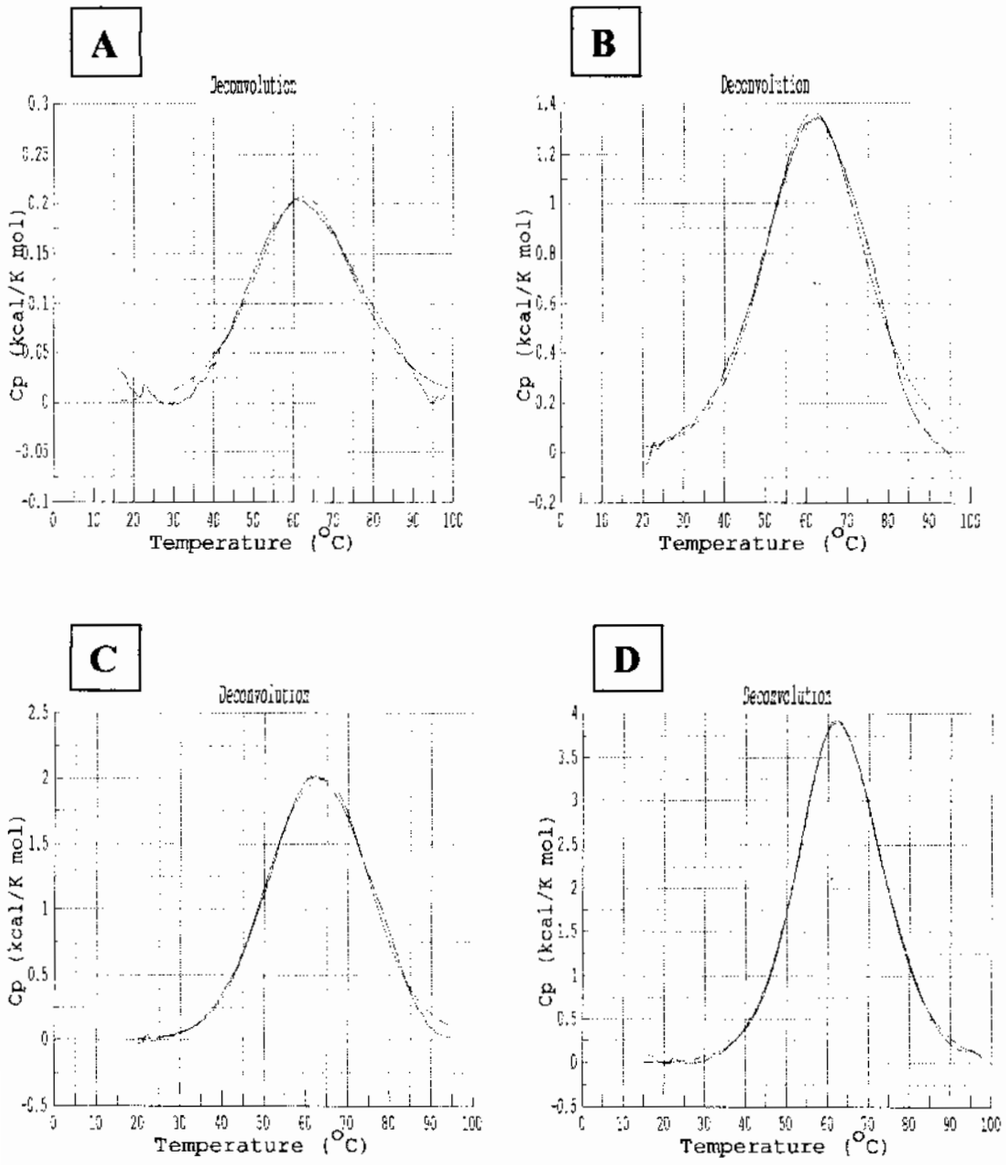


Figure 78: Representative van't Hoff modeling of the thermograms for the (CCG)n oligomers.

Representative "Two-State Scaled" data models for the (CCG)n series of oligomers, where n = 4, 10, 15, or 25 are shown below. For all data sets, the blue line represents the calorimetric data, and the red line represents the sum of the van't Hoff modeled data. **A.** The CCG4 oligomer. **B.** The CCG10 oligomer. **C.** The CCG15 oligomer. **D.** The CCG25 oligomer.

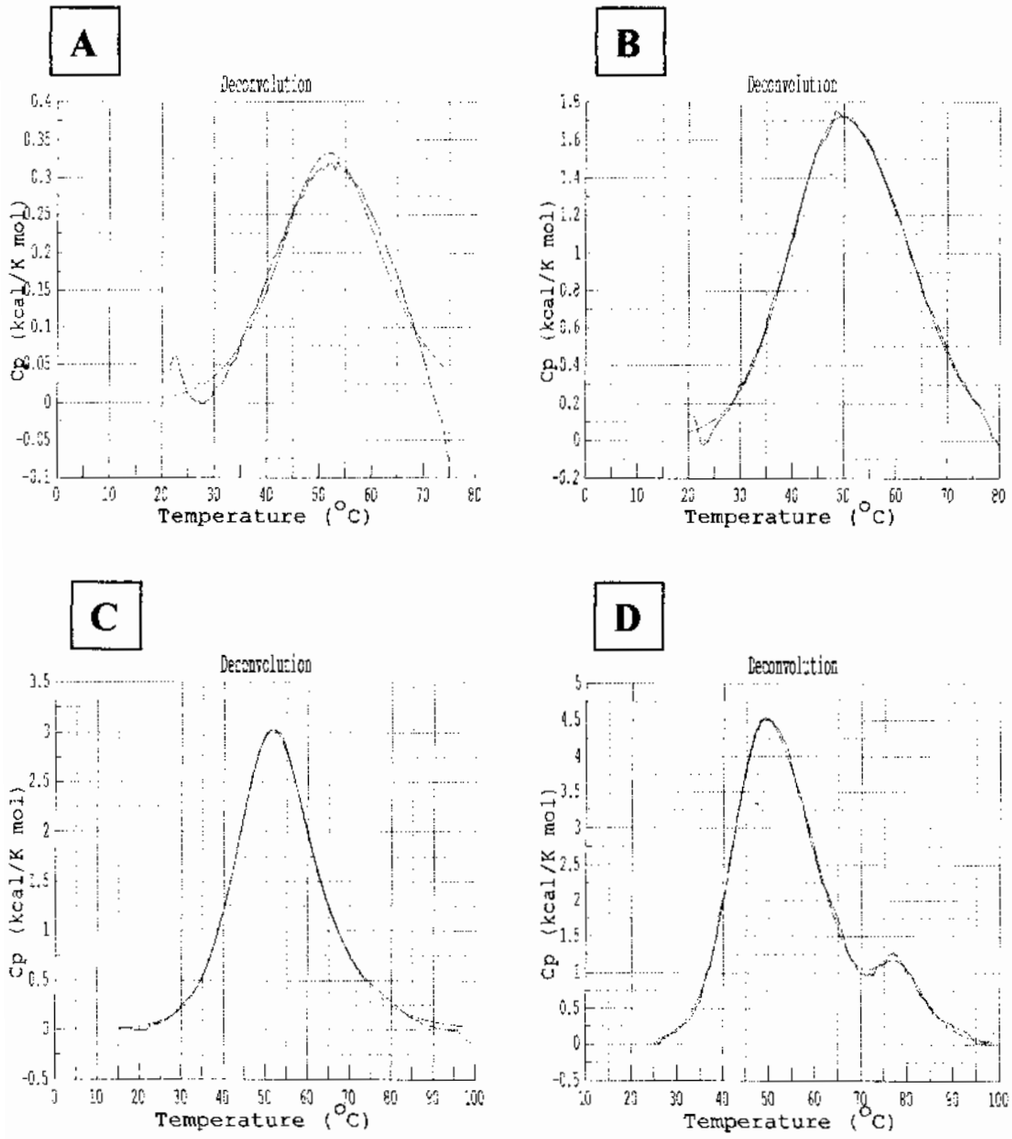


Figure 79: Representative van't Hoff modeling of the thermograms for the (CGG)n oligomers.

Representative "Two-State Scaled" data models for the (CGG)n series of oligomers, where n = 4, 10, 15, or 25 are shown below. For all data sets, the blue line represents the calorimetric data, and the red line represents the sum of the van't Hoff modeled data. **A.** The CGG4 oligomer. **B.** The CGG10 oligomer. **C.** The CGG15 oligomer. **D.** The CGG25 oligomer.

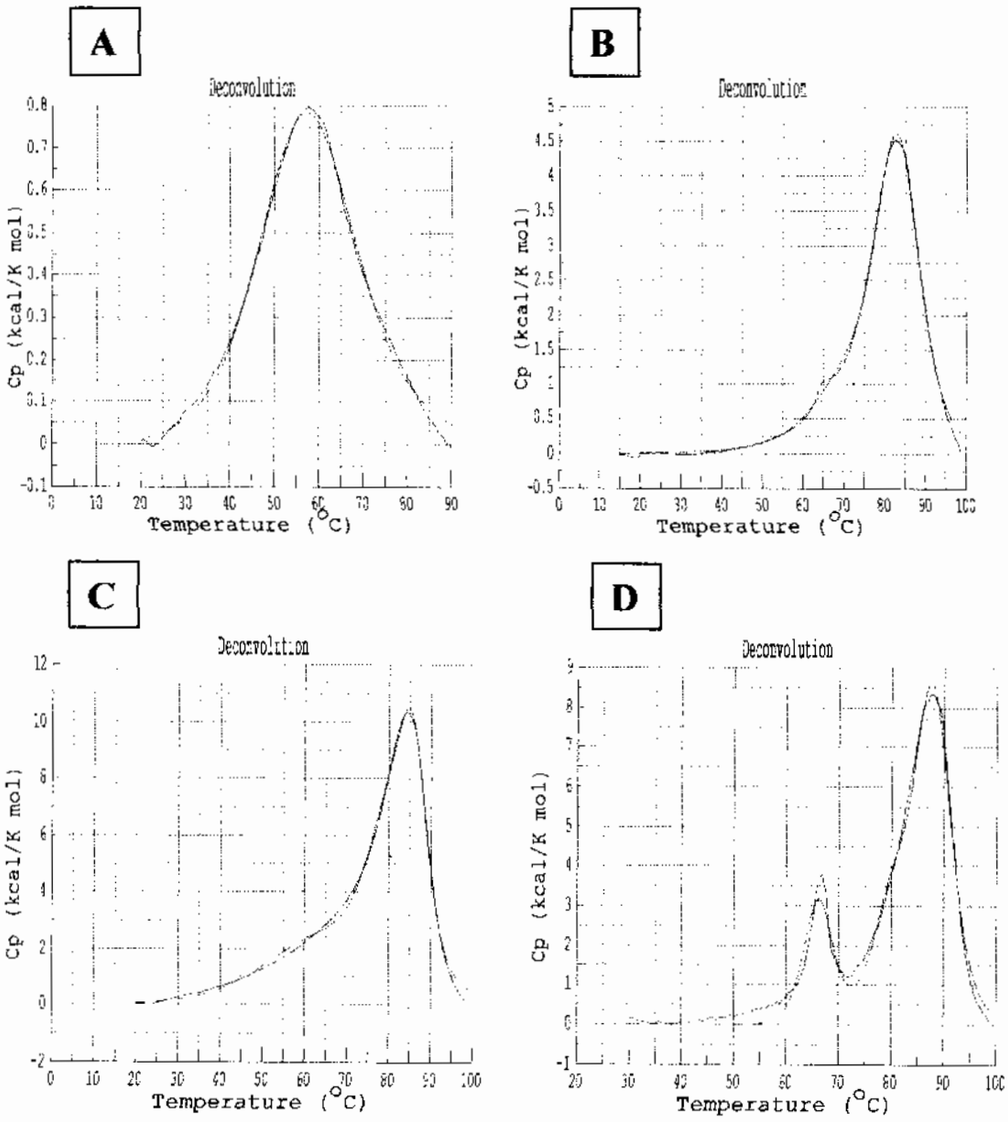
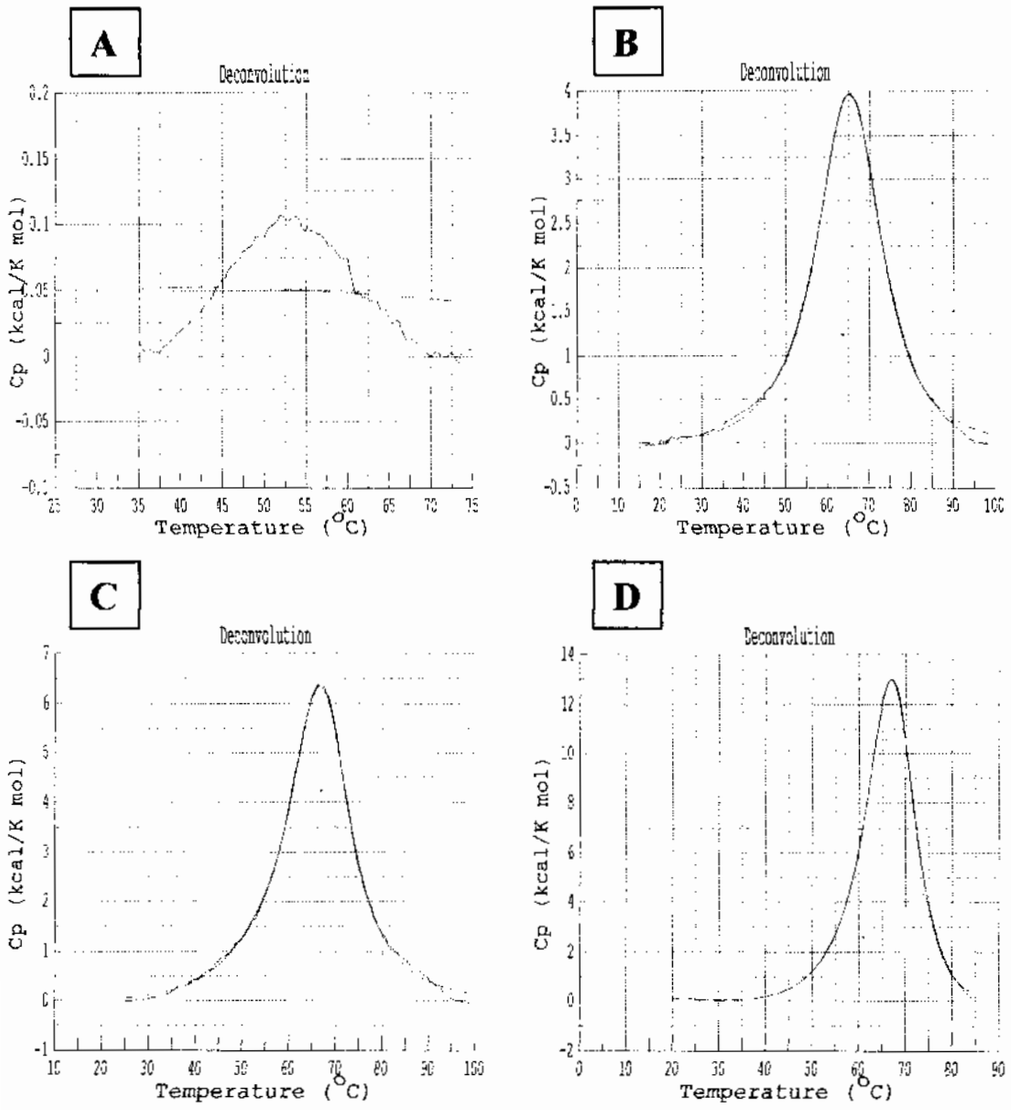


Figure 80: Representative van't Hoff modeling of the thermograms for the (CTG)n oligomers.

Representative "Two-State Scaled" data models for the (CTG)n series of oligomers, where n = 4, 10, 15, or 25 are shown below. For all data sets, the blue line represents the calorimetric data, and the red line represents the sum of the van't Hoff modeled data. **A.** The CTG4 oligomer, showing that no fit was possible. **B.** The CTG10 oligomer. **C.** The CTG15 oligomer. **D.** The CTG25 oligomer.



increased complexity of the peak shapes to be modeled increased the number of transitions required to model the data (Table 31). Hence, the longer the oligomer, the more transitions occurred during the melting process, and the less likely the melting transition resembled that of a true all-or-nothing interaction.

The modeled van't Hoff enthalpies for the (CAG)n oligomers, where n = 4, 10, 15, or 25 , increased from 8.0 kcal/mol for the 12-mer to 114 kcal/mol for the 75-mer (Table 31, Figure 77). The modeled enthalpy, based on calorimetric data, was inconsistent with the UV-melt van't Hoff enthalpy in terms of the magnitude due perhaps to an experimental bias in the UV-melt data. As previously described, short oligomers were fit with fewer transitions than longer oligomers due, in part, to the shape and complexity of the peaks modeled (Table 31). Comparison of the calorimetric enthalpy to the modeled calorimetric enthalpy revealed that most of the oligomers had ΔHvh/ΔHcal ratios that were either greater than, or significantly less than unity (Table 32). Comparison of the UV-melt derived enthalpy (modeled) to the calorimetric enthalpy revealed that the ΔHvh/ΔHcal ratios for all oligomers beyond the 12-mer were significantly lower than unity. These results suggest that that the (CAG)n oligomers did not melt in a true two-state fashion, but rather as a group of subpopulations.

Examination of the modeled van't Hoff enthalpies of the calorimetric data for (CCG)n and (CTG)n oligomers, where n = 4, 10, 15, or 25, displayed similar results with respect to the trend of the modeled calorimetric data. The modeled enthalpies increased as function of length, beginning from 49 kcal/mol and 60 kcal/mol

Table 31 : The van't Hoff transition enthalpy based on the calorimetric data.

Comparison of the melting transitions, and the van't Hoff enthalpies for each transition based upon the "Two-State Scaled" model of the calorimetric data. In the table below the ΔH_{vh1} corresponds to T_{m1}. All enthalpies are reported in kcal/mol.

Oligomer	T_{m1}	T_{m2}	T_{m3}	ΔH_{vH1}	ΔH_{vH2}	ΔH_{vH3}	$\Delta H_{vHtotal}$
CAG4	64.9			8.0	0.0	0.0	8.0
CAG10	58.7	69.7		30.3	26.6	0.0	57.0
CAG15	57.5	65.8	83.4	34.1	23.1	9.7	66.9
CAG25	55.4	62.8	71.4	22.6	52.7	38.6	114.0
Oligomer	T_{m1}	T_{m2}	T_{m3}	ΔH_{vH1}	ΔH_{vH2}	ΔH_{vH3}	$\Delta H_{vHtotal}$
CCG4	52.4			9.1	0.0	0.0	9.1
CCG10	47.7	58.7		31.4	17.6	0.0	49.1
CCG15	49.1	56.6	61.2	27.0	32.2	21.3	80.5
CCG25	46.5	55.7	76.5	55.5	64.1	18.4	138.0
Oligomer	T_{m1}	T_{m2}	T_{m3}	ΔH_{vH1}	ΔH_{vH2}	ΔH_{vH3}	$\Delta H_{vHtotal}$
CGG4	58.2			23.6	0.0	0.0	23.6
CGG10	73.6	83.2	72.5	26.6	34.0	5.8	66.4
CGG15	62.8	75.7	84.9	57.3	53.5	97.0	207.8
CGG25	65.5	82.0	88.5	49.9	52.0	70.3	172.2
Oligomer	T_{m1}	T_{m2}	T_{m3}	ΔH_{vH1}	ΔH_{vH2}	ΔH_{vH3}	$\Delta H_{vHtotal}$
CTG4				0.0	0.0	0.0	0.0
CTG10	64.0	65.2		25.3	34.6	0.0	60.0
CTG15	62.2	67.2		57.7	91.8	0.0	149.5
CTG25	62.8	67.4		78.1	106.5	0.0	184.6
Oligomer	T_{m1}	T_{m2}	T_{m3}	ΔH_{vH1}	ΔH_{vH2}	ΔH_{vH3}	$\Delta H_{vHtotal}$
CAG4/CTG4	59.9	61.0		22.9	30.8	0.0	53.7
CAG10/CTG10	67.9	79.5	84.8	32.7	56.1	47.8	136.6
CAG15/CTG15	75.2	84.4	88.0	56.7	99.6	59.2	215.5

Table 32: Comparison of the calorimetric and modeled data.

Comparison of the enthalpy ratios, $\Delta H_{vH}/\Delta H_{cal}$, between UV-melt and modeled calorimetric data. The original UV-melt enthalpy data was calculated for the SS to HP transition. The calorimetric data was calculated for the HP to SS transition. To ensure agreement between the two data sets, the sign of the UV-melt enthalpies were reversed.

Oligomer	dH_{vHUVM}	ΔH_{vHCal}	dH_{Cal}	$\Delta H_{vHCal}/\Delta H_{Cal}$	$\Delta H_{vHUVM}/\Delta H_{Cal}$
CAG4	16.3	8.0	5.5	1.5	3.0
CAG10	26.3	57.0	44.7	1.3	0.6
CAG15	29.0	66.9	56.4	2.0	0.5
CAG25	40.6	114.0	114.0	0.6	0.4
Oligomer	dH_{vHUVM}	ΔH_{vHCal}	dH_{Cal}	$\Delta H_{vHCal}/\Delta H_{Cal}$	$\Delta H_{vHUVM}/\Delta H_{Cal}$
CCG4	no fit	9.1	5.7	1.6	no fit
CCG10	28.7	49.1	57.0	0.9	0.5
CCG15	30.0	80.5	76.6	1.1	0.4
CCG25	27.5	138.0	101.8	1.4	0.3
Oligomer	dH_{vHUVM}	ΔH_{vHCal}	dH_{Cal}	$\Delta H_{vHCal}/\Delta H_{Cal}$	$\Delta H_{vHUVM}/\Delta H_{Cal}$
CGG4	20.4	23.6	23.2	1.0	0.9
CGG10	81.2	66.4	92.8	0.7	0.9
CGG15	65.0	207.8	147.8	1.4	0.4
CGG25	27.7	172.2	148.5	1.2	0.2
Oligomer	dH_{vHUVM}	ΔH_{vHCal}	dH_{Cal}	$\Delta H_{vHCal}/\Delta H_{Cal}$	$\Delta H_{vHUVM}/\Delta H_{Cal}$
CTG4	no fit	no fit	1.9	no fit	no fit
CTG10	53.3	60.0	77.1	0.8	0.7
CTG15	47.1	149.5	126.0	1.2	0.4
CTG25	41.4	184.6	165.0	1.1	0.3
Oligomer	dH_{vHUVM}	ΔH_{vHCal}	dH_{Cal}	$\Delta H_{vHCal}/\Delta H_{Cal}$	$\Delta H_{vHUVM}/\Delta H_{Cal}$
CAG4/CTG4	70.3	53.7	52.2	1.0	1.3
CAG10/CTG10	46.7	136.6	138.1	1.0	0.3
CAG15/CTG15	41.4	215.5	220.4	1.0	0.2
CAG25/CTG25	29.2	no fit	no fit	no fit	no fit

for the CCG10 and CTG 10 oligomers, respectively (Table 32). The modeled enthalpy, based upon the calorimetric data, then increased to 138 kcal/mol and 184 kcal/mol for CCG25 and CTG25 respectively. The observed ratios for the modeled calorimetric data to the calorimetric data suggested that the oligomers underwent a transition, from the 30-mer to the 75-mer that was non-two state. Both sets of oligomers began, at the 30-mer with a ΔHvh/ΔHcal ratio less than one, but then this ratio increased to greater than one as the length approached 45 nucleotides or greater. These results suggest that the oligomers begin as a non-two-state interaction with significant intermediates, and then aggregate as the length of the stand increases (Table 32). Comparison of the ΔHvh/ΔHcal ratio for the UV-melt and calorimetric data indicated that, while the nature of the melting curves are non-two-state, the population of the intermediate states increased as a function of increasing length – a trend that is opposite that suggested by the calorimetric models.

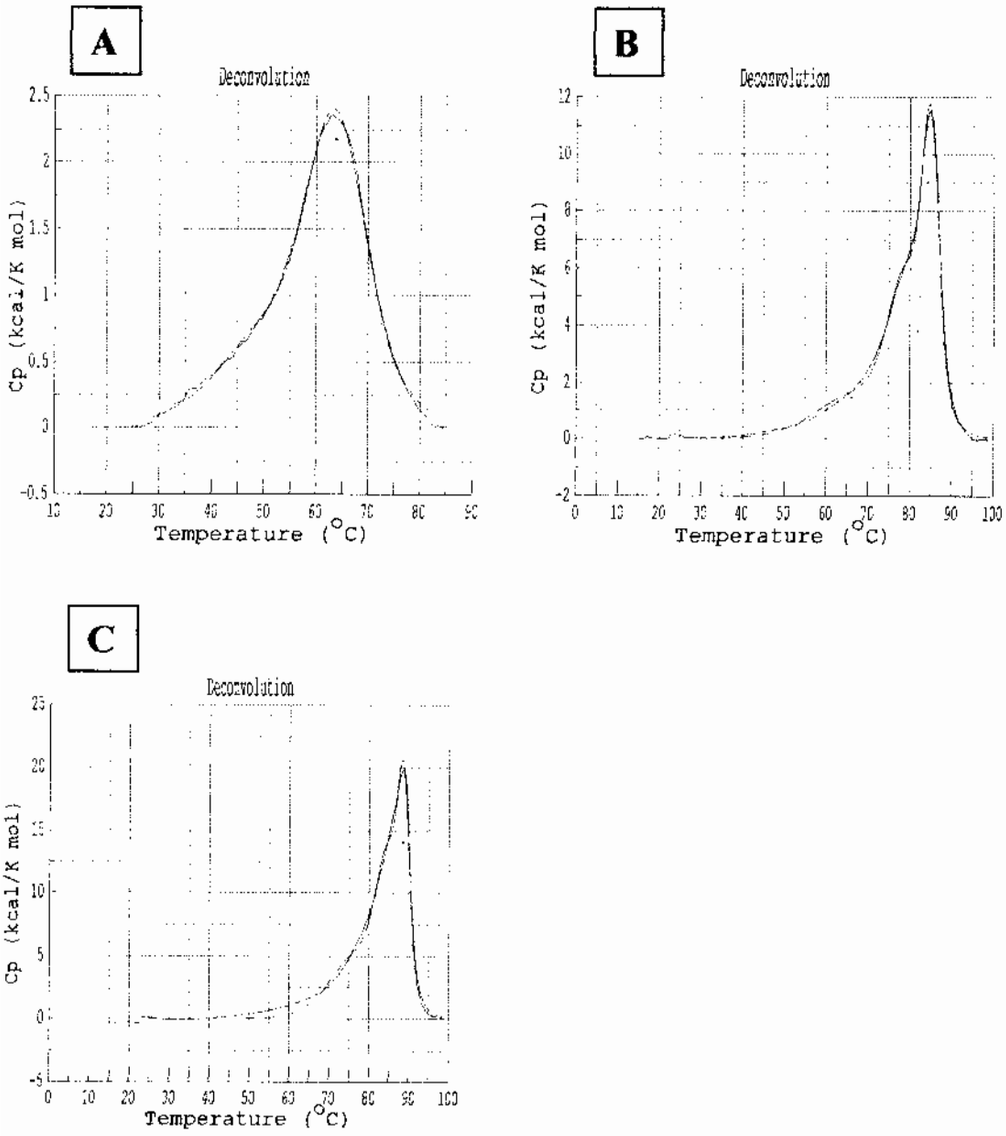
The van't Hoff enthalpies for the (CGG)n oligomers, where n = 4, 10, 15, or 25, displayed a peak in the enthalpic trend at approximately 45 nucleotides (Table 32). The modeled enthalpy increased from 24 kcal/mol for the 12-mer and peaked at 208 kcal/mol for the 45-mer. The modeled enthalpy then decreased, slightly, to 170 kcal/mol for the 75-mer. The observed trend, while mirroring the trend observed for the calorimetric enthalpy up to the 45-mer, did not resemble the calorimetric enthalpy trend as the length approached the 75-mer. The ΔHvh/ΔHcal ratio for the modeled calorimetric data followed similar trends that were observed for the (CCG)n and (CTG)n oligomers (Table 32). Namely the 12-mer, began as a two-state interaction that decayed into a non-two-state interaction with increasing length. Furthermore, the nature of the non-two-state

interactions underwent a transition, from a "simple" population of many intermediates for the 30-mer, with ΔHvh/ΔHcal less than one, to a possible aggregated state for the 45-mer and 75-mer, where ΔHvh/ΔHcal is greater than one.

Examination of the modeled van't Hoff enthalpies of the calorimetric data for the (CAG)n/(CTG)n duplexes, where n = 4, 10, or 15, displayed magnitudes that were similar to those obtained by calorimetry (Figure 81, Table 32). The enthalpic trend increased as a function of increasing duplex length, from 54 kcal/mol for the 12-bp duplex to 215 kcal/mol for the 45-bp duplex. The $\Delta H_{vH}/\Delta H_{cal}$ ratio between the modeled calorimetric and calorimetric data was one for all duplex lengths studied within the (CAG)n/(CTG)n duplex series. This finding suggests that the duplexes underwent true two-state (all-or-nothing) type of melting transitions. This data indicates a fundamental difference in the thermodynamic properties between the solitary oligomers that are capable of forming mis-matched hairpins, and these same sequences that form complementary duplexes. However, the $\Delta H_{vH}/\Delta H_{cal}$ ratio between the UV-melt data and the calorimetric data was inconsistent with this conclusion, and underscored the inaccuracies and/or difficulties that were associated with obtaining van't Hoff data from UV-melt curves.

Figure 81: Representative van't Hoff modeling of the thermograms for the (CAG)n/(CTG)n duplexes.

Representative "Two-State Scaled" data models for the (CAG)n/(CTG)n duplexes, where n = 4, 10, 15, or 25 are shown below. For all data sets, the blue line represents the calorimetric data, and the red line represents the sum of the van't Hoff modeled data. **A.** The CTG4 oligomer. **B.** The CTG10 oligomer. **C.** The CTG15 oligomer. **D.** The CTG25 oligomer.



4.4 Kinetics

4.4.1 Arrhenius Activation Energy Plots

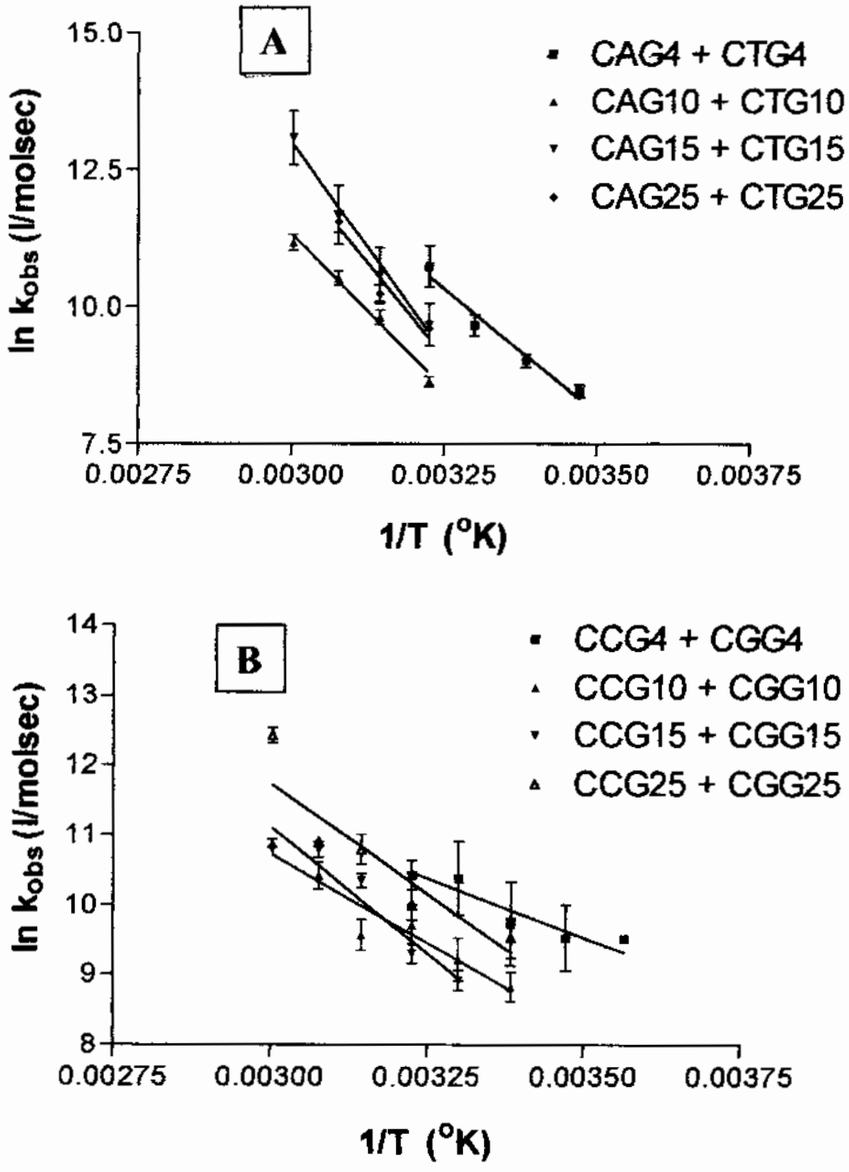
4.4.1.1 Sequence and Length-Dependent Activation Energy Trends.

The work of Gacy and McMurray indicated that there was a length-dependent difference with respect to the reassociation rate constants for the HP to DUP association process.[29] This observation suggested a possible kinetic role in the association of two complementary hairpins and perhaps a relationship to the energy of activation that is length dependent. To obtain the energy of activation the rate constants of the HP to DUP association process were measured for each complementary (CNG)n/(CNG)n oligomer pairs, where N/N = A/T or C/G, and n = 4, 10, 15 or 25, as a function of increasing temperature (Figure 82). The reassociation rates were measured in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH7.0, 100 mM NaCl at temperatures that ranged from 7.5 °C for the lowest point of the 12-mers, to 60 °C, the highest data point for the 75-mers. For most oligomer sets, the upper and lower limits of the temperature range varied, depending upon the quality of the acquired data. At elevated temperatures, beyond 42 °C, for example, the observed rate constants for the 12-mers were too fast for accurate quantitation. Likewise the lowest temperature that produced acceptable rate constants for the 12-mer, was found to be too low for accurate measurement of the 75-mer rate constant.

The observed rate constants with respect to sequence context revealed that overall, the (CAG)n/(CTG)n complementary oligomers reassociated to duplex with a higher energy

Figure 82: Arrhenius plots for the HP to DUP association process.

Arrhenius plots were used to derive the energy of activation for the HP to DUP association process. The plots indicate, from the slope of the lines, that the (CAG)n/(CTG)n oligomers have higher activation energies than do the (CCG)n/(CGG)n oligomers. In addition, the plots also indicate, based on the slopes of the lines that short oligomers have lower activation energies for the HP to DUP reassociation process than do long oligomers. **A.** The Arrhenius activation energy plot for the (CAG)n and (CTG)n oligomers. **B.** The Arrhenius activation energy plot for the (CCG)n and (CGG)n oligomers.



of activation, and slower kinetics than did the (CCG)n/(CGG)n oligomer pairs (Table 33).

Table 33: Arrheniuis activation parameters for the HP to DUP association process.

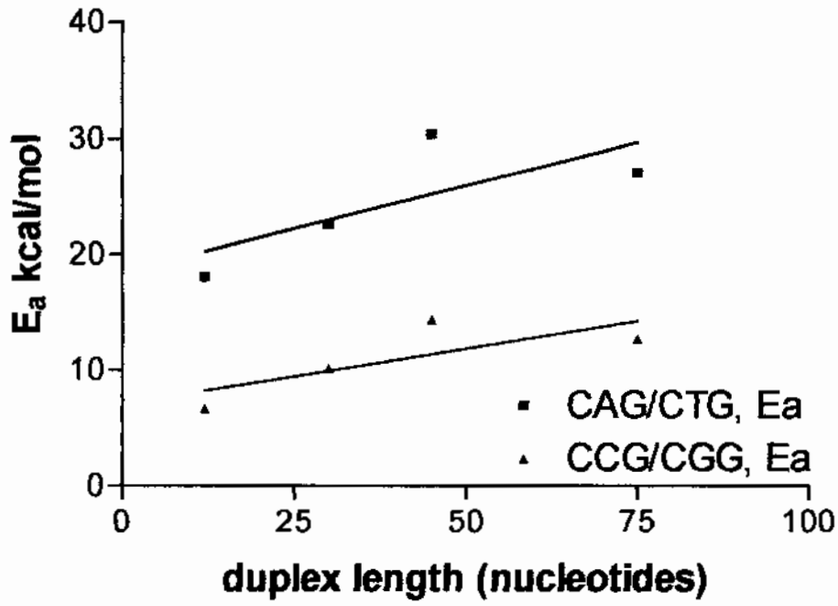
Oligomer Pair	Slope	Intercept	r^2	E_a
CAG4/CTG4	-9085	39.86	0.79	18.1
CAG10/CTG10	-11310	45.28	0.93	22.5
CAG15/CTG15	-15290	58.86	0.75	30.4
CAG25/CTG25	-13590	53.26	0.84	27.0
Oligomer Pair	**Slope**	**Intercept**	r^2	E_a
CCG4/CGG4	-3382	25.47	0.22	6.7
CCG10/CGG10	-5062	30.1	0.68	10.1
CCG15/CGG15	-7215	32.14	0.85	14.3
CCG25/CGG25	-6317	30.69	0.74	12.6

On average the energy of activation for the (CAG)n/(CTG)n oligomer pairs were approximately 2-fold higher than the energy of activation for the (CCG)n/(CGG)n oligomer pairs (Figure 82). Due to the similarity of the triplet repeat sequences, where only the identity of the middle or mismatched nucleotide is varied, the data suggest that there is a sequence-context effect with respect to strand reannealing. Were there no such sequence context effects, the observed rate constants and the Arrhenius energy of activation parameters would have been similar between the two data sets. This suggests a possible reason for the observed difference in expansion frequencies for the CAG series of oligomers. The reannealing rates of the (CAG)n/(CTG)n hairpin to duplex association are slower than those of the corresponding (CCG)n/(CGG)n sequences.

As a function of length, the Arrhenius energy of activation for the (CAG)n/(CTG)n oligomers began from 18 kcal/mol for the 12-mer association, and peaked at 30 kcal/mol for the 45-mer association (Figure 83, Table 33). The energy of activation, then decreased to 27 kcal/mol for the 75-mer association process. On the other hand, the (CCG)n/(CGG)n oligomers displayed a similar, although lower energy of activation trend. The activation energy began from 7 kcal/mol for the 12-mer association process and then peaked at 15 kcal/mol for the association of the 45-mers. The energy of activation then decreased by approximately 2 kcal/mol, to 12.7 kcal/mol, for the association of the 75-mers CCG25 with CGG25. On average, for every base-pair added to the duplex, the activation energy for the (CAG)n/(CTG)n oligomer pairs increased by approximately 0.36 kcal/mol basepair, versus 0.17 kcal/mol basepair for the (CCG)n/(CGG)n duplexes. When the data were fit with linear regression analysis (Figure 83) a similar trend in the slope of the activation energy versus length plot was obtained – the (CAG)n/(CTG)n oligomer pairs had higher activation energies than did the (CCG)n/(CCGG)n oligomer pairs. The difference in slope between the two data sets was approximately 1.6-fold – for every basepair added to the duplex, the linear regression increased by 1.60 fold more for the CAG/CTG oligomers. The data trends indicated length dependence with respect to the energy of activation. The longer the segment of DNS becomes, the slower the HP reanneals to the corresponding duplex. The observed trend therefore suggests that there is s kinetic component with respect to disease causation – longer sequences triplet repeat sequences take longer to reanneal to the duplex than do short segments of triplet repeat DNA.

Figure 83: Arrhenius activation energy as function of length.

A plot of the activation energy as a function of length shows a length-dependent increase in the activation energy. In addition, the plot shows that the (CAG)n/(CTG)n oligomers have higher activation energies than do the (CCG)n/(CGG)n class of oligomers.



Variables	CAG/CTG	CCG/CGG
Slope	0.15	0.09
Y-intercept	18.50	7.08
r^2	0.54	0.58

4.4.2 Eyring Plots of the Kinetic Data

4.4.2.1 Enthalpy and Entropy of Activation Trends with Respect to Sequence Context and Length.

The kinetic data from the Arrhenius studies were replotted as an Eyring plot to determine the enthalpy and entropy of activation for the two sequence classes under consideration (Figure 84). In general, the Eyring plots revealed enthalpic and entropic trends that were similar to the observed trends for the energy of activation. The enthalpy of activation was generally higher for the (CAG)n/(CTG)n oligomer pairs than it was for the (CCG)n/(CGG)n oligomer pairs (Table 34). Overall, the enthalpies of activation for the

Table 34: Eyring plot parameters for the HP to DUP association process.

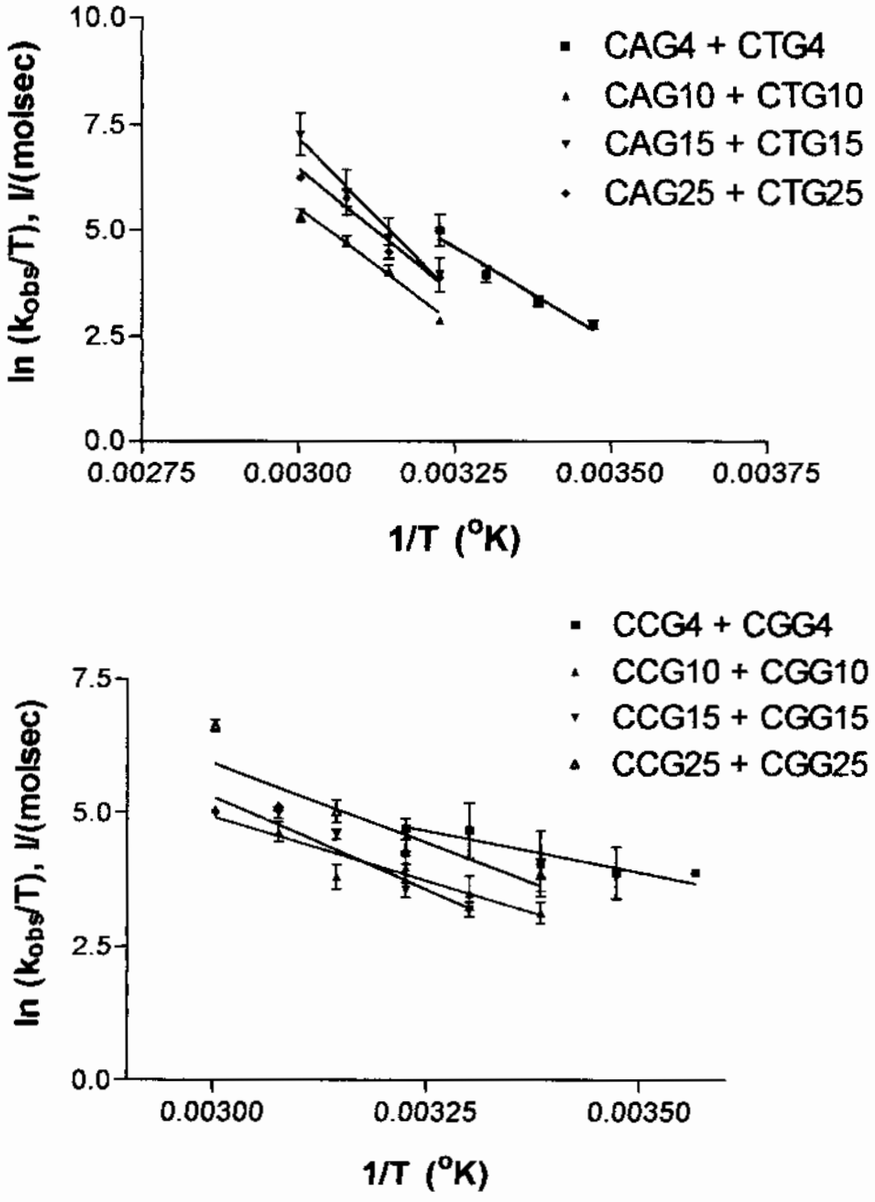
The Eyring plot parameters for the transition state enthalpy, in kcal/mol, and transition state entropy, in cal/(mol K), are shown below.

Oligomer Pair	Slope	Intercept	r^2	ΔH_{ts}	ΔS_{ts}
CAG4/CTG4	-8790	33.17	0.78	17.5	18.7
CAG10/CTG10	-11000	38.54	0.92	21.9	29.4
CAG15/CTG15	-14970	52.1	0.74	29.7	56.3
CAG25/CTG25	-11890	42.13	0.87	23.6	36.5
Oligomer Pair	**Slope**	**Intercept**	r^2	ΔH_{ts}	ΔS_{ts}
CCG4/CGG4	-3078	14.66	0.2	6.1	-18.1
CCG10/CGG10	-4746	19.15	0.65	9.4	-9.2
CCG15/CGG15	-6901	26.01	0.84	13.7	4.5
CCG25/CGG25	-6000	23.93	0.72	11.9	0.3

CAG/CTG oligomer pairs were approximately 2-fold higher than the enthalpies of activation for the CCG/CGG oligomer pairs. In a similar manner, the entropies of

Figure 84: Eyring plots of the HP to DUP association process.

Eyring plots were used to derive the enthalpic and entropic contributions to the transition state kinetics. **A.** The Eyring plot for the (CAG)n and (CTG)n oligomers. **B.** The Eyring plot for the (CCG)n and (CGG)n oligomers.



activation for the (CAG)n/(CTG)n oligomers were higher, and indicated more disorder in the transition state of the (CAG)n/(CTG)n oligomer pairs than with the (CCG)/(CGG)n oligomer pairs. The data therefore suggest that the (CAG)n/(CTG)n oligomers undergo extensive bond breaking, with more disorder in the transition state than do the (CCG)n/(CGG)n oligomer pairs.

When a comparison of length was undertaken, a trend similar to the trend observed for the energy of activation was found (Table 34). The enthalpy of activation became larger as the length of the oligomers approached the 30-mer indicating more bond breakage as a function of length. The enthalpy of activation began from approximately 17 kcal/mol and 6.1 kcal/mol for the CAG4/CTG and CCG4/CGG4 oligomer pairs respectively, and peaked at approximately 30 kcal/mol and 14 kcal/mol respectively for the CAG10/CTG10 and CCG10/CGG10 oligomer pairs. At the same time, the entropy of activation became larger, indicating more disorder as the oligomers approached the transition state up to the 30-mer. Beyond this length the transition state enthalpy and entropy trends displayed a slight decrease in magnitude and was suggested a plateau or leveling-off of the data trend.

Overall, the replotted data from the Arrhenius experiments suggested a possible explantion for the observed difference in the activation energies between the (CAG)n/(CTG)n and (CCG)n/(CGG)n oligomer pairs. The (CAG)n/(CTG)n oligomers undergo more bond breaking, as shown by the higher enthalpy term, and and have more disorder in the transition state, as shown by the higher entropy term, than the

(CCG)n/(CGG)n oligomer pairs. Together enthalpy and entropy decrease the probability that the single-stranded end, once formed, will find a complementary segment of DNA with which to nucleate and anneal. The (CCG)n/(CGG)n oligomer pairs, on the other hand have less disorder, and therefore a more structured transition state that encourages nucleation.

4.4.3 Sodium Ion Effects – A Test of the Mechanistic Pathway

4.4.3.1 Sodium Ion Effects

The hairpin form of triplet repeat DNA is a thermodynamically unstable, or metastable state, that will eventually anneal to a complementary strand and form a duplex DNA structure. The mechanism through which DNA hairpins can undergo this change in structure can follow two different intermediates, according to the theoretical framework of Azivonas and Kearns.[38] One of these intermediates, dubbed the single-stranded intermediate, is characterized by high amounts of hydrogen-bond breakage and high energies of activation (Figure 37). Because the intermediate in this mechanism is single-stranded, the reaction kinetics will be sensitive to the effects of Na^+ concentration. In general, increasing Na^+ concentration destabilizes the single-stranded DNA transition state and increases the kinetics. The other intermediate is the cruciform (CR) that is characterized by low amounts of bond breakage and low energies of activation (Figure 37). Unlike the single-strand transition state, the cruciform intermediate is stabilized by increasing Na^+ concentration. Therefore, increasing the Na^+ concentration increases the rate of the reverse, CR to HP conversion, with the net effect of slowing the observable

reaction kinetics (Figure 37). Up to this point in the kinetic characterization of the triplet repeat sequences, the following information was obtained. First, the energy of activation, the enthalpy of transition and the entropy of transition are higher for the CAG/CTG oligomers than it is for the CCG/CGG oligomers. This observation suggests that the two sequence classes undergo fundamentally different mechanisms as the hairpins or other structure nucleate and then anneal to form the duplex. With respect to the CAG/CTG oligomer pairs, the transition state involves a large amount of bond breaking, and, relative to the CCG/CGG oligomer pairs, a larger amount of disorder. Similarly, as the length of the oligomer sequence, or triplet repeat region increased, there was an observed an increase in the enthalpy (the number of bonds broken) and entropy (more disorder) as a function of increasing length. Taken together, these facts suggest the possibility that the two classes of oligomer sequence pairs, CAG/CTG, and CCG/CGG undergo fundamentally different mechanisms in the reanimating process. The data suggested that perhaps the CAG/CTG strands underwent a transition state that was more single-stranded in nature relative to the CCG/CGG oligomer pairs. With these facts in mind, an attempt to verify the mechanism of the hairpin to duplex mechanism was made by studying the effects of Na^+ concentration on the reaction kinetics.

The HP to DUP association kinetics of the (CNG)n oligomer pairs, where N = A/T or C/G and n = 4, 10, 15, or 25, were studied in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH 7.0 with sodium chloride added to achieve a final concentration of 15, 25, 50, 100, 250, 500 mM Na^+. For all reactions, the solitary oligomers were equilibrated and the association reaction measured at 37 °C. Overall, the (CAG)n/(CTG)n oligomers show

little difference in observed rate constants when compared to the (CCG)n/(CGG)n oligomer pairs (Figure 85, Table 35). Linear regression analysis of the observed rates

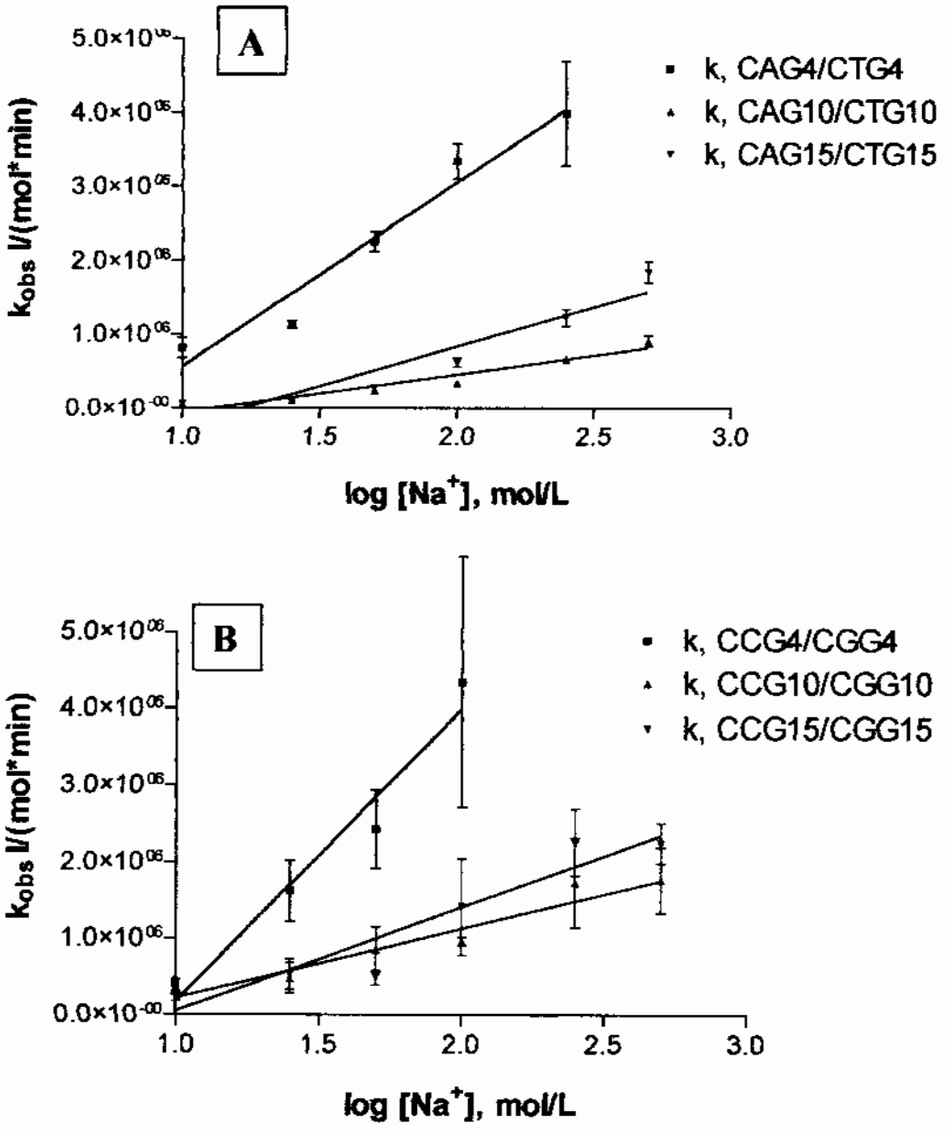
Table 35 : Linear regression parameters of the effect of Na^+ concentration on the observed rate constants.

Variables	CAG4/CTG4	CAG10/CTG10	CAG15/CTG15
Slope	2490000	510200	1065000
Y-intercept	-1926000	-563300	-1300000
r^2	0.77	0.89	0.86
Variables	**CCG4/CGG4**	**CCG10/CGG10**	**CCG15/CGG15**
Slope	3819000	893900	1347000
Y-intercept	-3637000	-666800	-1293000
r^2	0.42	0.52	0.66

constants for the 12-mer association reaction as a function of Na^+ concentration showed approximately a two-fold difference between the (CAG)n/(CTG)n and (CCG)n/(CGG)n oligomer pairs. However, these the observed trend did not continue with the longer sequences, such as the 30-mer or 45-mer association reactions. When examined for a length dependence, both the (CAG)n/(CTG)n and (CCG)n/(CGG)n oligomer pairs displayed rates, as a function of Na^+ concentration that were consistently higher for the 12-mers than for the corresponding 30-mer or 45-mer. This indicates that the 12-mers have relatively higher rate constants in response to Na^+ concentration, and suggests that the 12-mer transition state is more single-strand-like than the transition state for the 30-mer or 45-mer (Table 35).

Figure 85: Efect of sodium ion concentration on the observed rate constants.

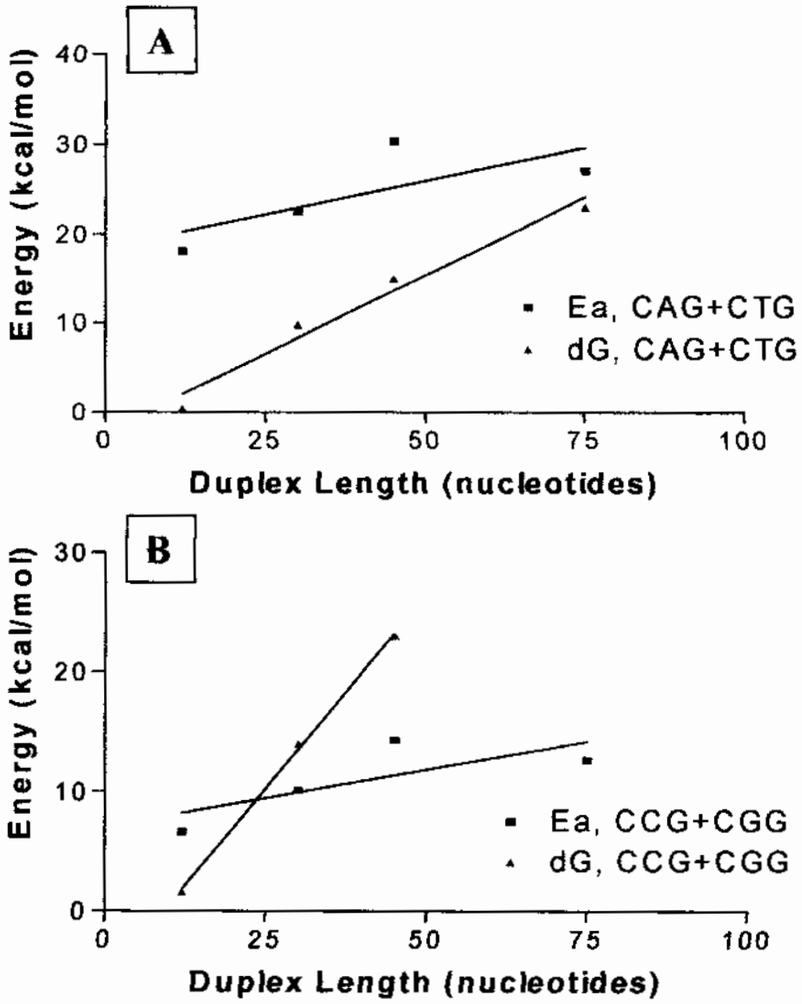
Plots of the effect of Na^+ concentration on the observed rate constants indicate that there is a difference in the kinetic behavior of the short hairpins (12-mer) than for the long hairpins (30-mer and 45-mer). **A.** The Na^+ concentration plot for the (CAG)n and (CTG)n oligomers. **B.** The Na^+ concentration plot plot for the (CCG)n and (CGG)n oligomers.



When viewed from a slightly different perspective, the work of Azivonas and Kearns sheds additional light on the nature of the transition state, and the underlying reasons behind the observed differences in rate constants for the reannealing of the (CAG)n/(CTG)n and (CCG)n/(CGG)n oligomer pairs. An additional test of the nature of the transition state is to examine the relationship between the energy of activation and the thermodynamic stability of the strands under consideration. As previously described, a single-stranded intermediate requires the simultaneous breaking of multiple hydrogen bonds. Bond-breakage on such a scale involves enormous amounts of energy, such that the energy of activation will equal to or exceed the thermodynamic stability of the DNA structure under consideration (Figure 37).[38] A cruciform intermediate however, requires the breakage of only a few hydrogen bonds at a time and reduces the amount of energy required to accomplish the reaction. For such a processs the energy of activation is below the thermodynamic stability of the DNA structure under consideration (Figure 37).[38] A plot of the energy of activation, derived from the Arrhenius data at 115 mM Na^+, and the sum of the calorimetric thermodynamic stabilities, derived from ΔG_{cal}, for each hairpin oligomer under consideration revealed several trends (Figure 86). First, with respect to the (CAG)n/(CTG)n oligomer pairs, the activation energy was higher than the calorimetric thermodynamic stability for almost all lengths of oligomer pairs studied. Extrapolation of the trend-lines indicated that the summed thermodynamic stability for the (CAG)n/(CTG)n oligomer pairs did not become larger than the energy of activation until approximately 100 base-pairs in length. In addition, the difference between the thermodynamic stability and the energy of activation became smaller as the length of the (CAG)n/(CTG)n oligomer pair increased. With respect to the (CCG)n/(CGG)n oligomer

Figure 86: Comparison of Ea and ΔG of stability for the HP to SS transition .

A comparison between the summed thermodynamic stability of each hairpin under consideration ΔG, and the activation energy, Ea, for the association of two complementary hairpins is shown below. **A.** A comparison of the (CAG)n, (CTG)*n* hairpin stabilities versus the energy of activation indicates that the Ea is higher than the thermodynaic stbiality. **B.** A comparison of the (CCG)n and (CGG)n hairpin stabilities with the energy of activation indicates that the hairpin association process undergoes a single-stranded intermediate for short sequences, but then a cruciform-like intermediate as the length increases.



pairs, the activation energy 12-mer oligomer pairs were higher than the summed thermodynamic stability for the individual 12-mers. There was a cross-over in the trend lines at approximately 25-bp and by 30-bp where the summed thermodynamic stability surpassed the energy of activation. In a general view, these observations suggest that the (CAG)n/(CTG)n oligomers undergo fundamentally different transition states and/or pathways than do the (CCG)n/(CGG)n oligomers when reannealing is under consideration. The data suggest that the (CAG)n/(CTG)n oligomer pairs undergo a more single-strand-like transition state. This is consistent with the observations from the Eyring plots that suggested the (CAG)n/(CTG)n oligomers undergo more hydrogen-bond breakage and have more disorder in the transition state. In addition, as the length of the (CAG)n/(CTG)n strands increase the transition state becomes less single-stranded in nature and more cruciform-like in nature. The data also suggest that the (CCG)n/(CGG)n oligomer pairs undergo a more cruciform-strand-like transition state. This observation is consistent with the data from the Eyring plots that suggested the (CCG)n/(CGG)n oligomers undergo less hydrogen-bond breakage and have more order in the transition state than do the (CAG)n/(CTG)n oligomers. However, in contrast to these findings is the observation from the same Eyring plot data that indicate the entropy of the transition state becomes larger, as a function of length (Figure 85 vs. Table 34). It is possible therefore that the nucleation site takes on a single-stranded/cruciform behavior, the overall trend of the molecule is for increasing entropy due to the increased conformational freedom associated with increased size.

5 DISCUSSION/CONCLUSIONS

5.1 Structural Studies

5.1.1 Purpose

There has been much discussion about the nature of the structures formed by triple repeat DNA sequences. Many researchers have observed these sequences to form hairpin structures, on the basis of gel electrophoresis,[16] or quadruplexes on the basis of solution NMR measurements.[15] Because of the diverse solution conditions employed by the different research laboratories, and the lack of a complete database encompassing both sequence and length comparison we undertook the task of characterizing the structures formed by triplet repeat DNA under physiological conditions. In addition, such studies prove to be critical for establishing the nature of the DNA structure prior to structural perturbation studies such as melting.

5.1.2 Solitary Oligomers form Hairpin Structures

In this dissertation we show, on the basis of CD spectra, that the solitary oligomers can form DNA structures with lengths as short as 12 nucleotides long (Figure 40, Figure 41). With the exception of the (CGG)n oligomers, all CD spectra bear a striking resemblance to each other, and to the idealized CD-spectra of B-form DNA. This dissertation also shows that the overall nature of these structures changes very little within a sequence class regardless of the length of the oligomer under consideration (Figure 40, Figure 41). While there are minor changes in the height of a peak or trough, indicative of changes in stacking interactions, or changes in the location of the maxima/minima indicating changes in hydration, the overall structures remain intact. The CD spectral data suggest

the oligomers are in a structural conformation similar to that of B-form DNA. Typically, B-form DNA has a peak centered at 280 nm, a crossover point centered near 260 nm and a trough centered at approximately 255 nm. The peak maxima, crossovers, and troughs for the (CAG)n, (CCG)n and (CTG)n oligomers were all centered at or near the B-form DNA benchmarks. As previously noted, the CD spectral signals were not exactly the same as B-form DNA due possibly to two reasons. First, triplet repeat DNA sequences have an unusually high GC content that ranges from 66 % G/C for CAG and CTG, to 100 % G/C for CCG or CGG. It has been shown by Johnson that DNA sequence content can alter the observed CD spectrum of a DNA molecule from that of an idealized B-form CD spectrum.[164] This phenomenon was related to the DNA sequence, and was attributable to the high sequence content of a particular nucleotide. The additive effects of a particular nucleotide in abundance can alter the CD spectrum. Second, the presence of mismatches located every third nucleotide can serve to alter the CD spectrum, by introducing structural and/or spectroscopic changes that affect the nucleotide interactions.

While the CD spectra indicate that the (CNG)n oligomers can form stable structures, there are two possible structures that can be formed from the DNA sequences under consideration – either a mismatched hairpin or a mismatched duplex, where every third nucleotide in the sequence is mismatched (Figure 88). According to the work of Zheng[15] and coworkers the (CNG)n sequences can also form, under the appropriate solution conditions, higher order structures such as quadruplexes. In this thesis we find, with the exception of the (CGG)n CD spectra, no unusual CD spectral properties associated with the (CNG)n oligomers that suggest higher order structures. Even with the possibility of a

Figure 87: Hairpin versus mismatched duplex form of the (CNG)n repeats.

The hairpin and duplex forms of the CAG10 oligomer are depicted below. In both structures, the oligomers contain a mismatch every third nucleotide shown by the lack of a heavy line between the mismatches. In the figure below, hydrogen bonding is represented schematically by a solid line between complementary base-pairs. The number of hydrogen bonds is 3 per G/C base pair. Both structures should produce similar CD spectra due to the similarity between the duplex and the stem of the hairpin. The loop region of the hairpin should not significantly contribute to the CD spectrum due to the absence of base-stacking interactions, and the small size of the loop relative to the stem region.

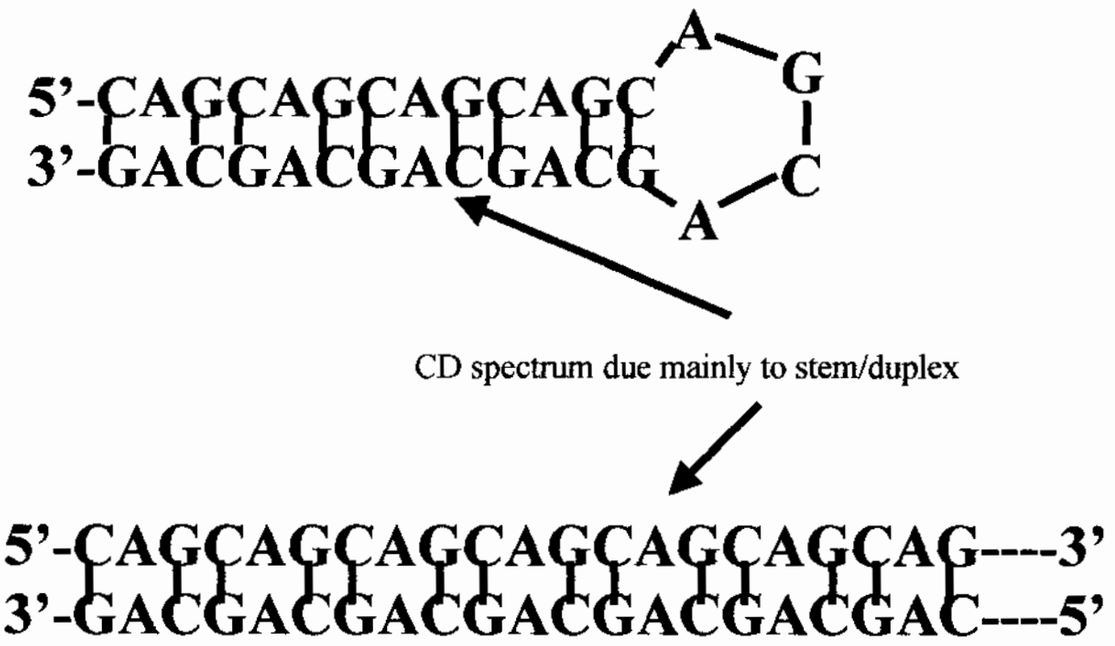
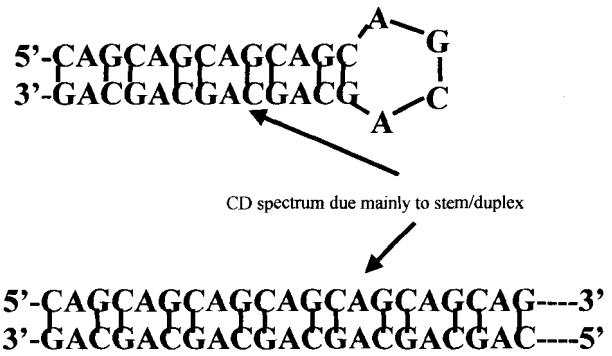


Figure 87: Hairpin versus mismatched duplex form of the (CNG)n repeats.

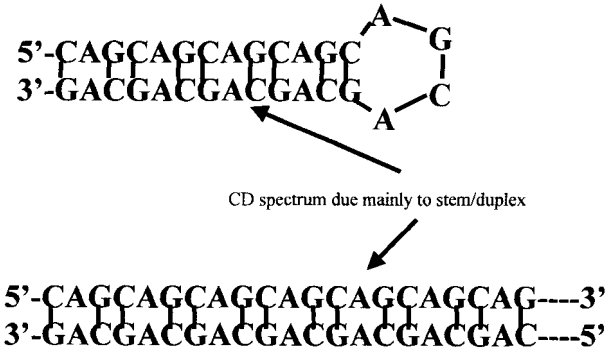
The hairpin and duplex forms of the CAG10 oligomer are depicted below. In both structures, the oligomers contain a mismatch every third nucleotide shown by the lack of a heavy line between the mismatches. In the figure below, hydrogen bonding is represented schematically by a solid line between complementary base-pairs. The number of hydrogen bonds is 3 per G/C base pair. Both structures should produce similar CD spectra due to the similarity between the duplex and the stem of the hairpin. The loop region of the hairpin should not significantly contribute to the CD spectrum due to the absence of base-stacking interactions, and the small size of the loop relative to the stem region.



CD spectrum due mainly to stem/duplex

Figure 87: Hairpin versus mismatched duplex form of the (CNG)n repeats.

The hairpin and duplex forms of the CAG10 oligomer are depicted below. In both structures, the oligomers contain a mismatch every third nucleotide shown by the lack of a heavy line between the mismatches. In the figure below, hydrogen bonding is represented schematically by a solid line between complementary base-pairs. The number of hydrogen bonds is 3 per G/C base pair. Both structures should produce similar CD spectra due to the similarity between the duplex and the stem of the hairpin. The loop region of the hairpin should not significantly contribute to the CD spectrum due to the absence of base-stacking interactions, and the small size of the loop relative to the stem region.



CD spectrum due mainly to stem/duplex

higher order species eliminated, at least for the (CAG)n, (CCG)n and (CTG)n oligomers, there still remained much ambiguity about the nature of the structures formed given the available CD spectral data. Both the hairpin and mismatched duplex structures should produce similar CD spectra due to the structural similarity between the duplex and the stem of the hairpin. The loop region of the hairpin, however, should not significantly contribute to the CD spectrum due to the absence of base stacking interactions, and the small size of the loop relative to the stem region. Due to the spectral similarity between these two structural classes of DNA, there can be no assignment of structure without ancillary evidence. For this reason, the concentration dependence of the melting transition and native gel electrophoresis was undertaken.

This dissertation demonstrates that the DNA concentration dependence of the melting transition for all (CNG)n oligomers, where n = 4, 10 or 15, (except CGG4 and CTG4) were nearly independent of DNA concentration. This indicates that all solitary oligomers formed structures that melted as unimolecular units. Were the oligomers forming mismatched duplexes the melting transition would have shown a net increase as the DNA concentration increased. Higher order DNA structures tend to increase the stability of the assembly, and increase the melting transition by decreasing the value of the equilibrium constant at the melting transition (Equation 1). Therefore, either a mismatched duplex with a molecularity of 2, or a quadruplex with a molecularity of 4 would have shown concentration dependence with a negative slope. In addition, a higher order structure such as a tetraplex would also have shown a multiple transition UV-melt curve, with

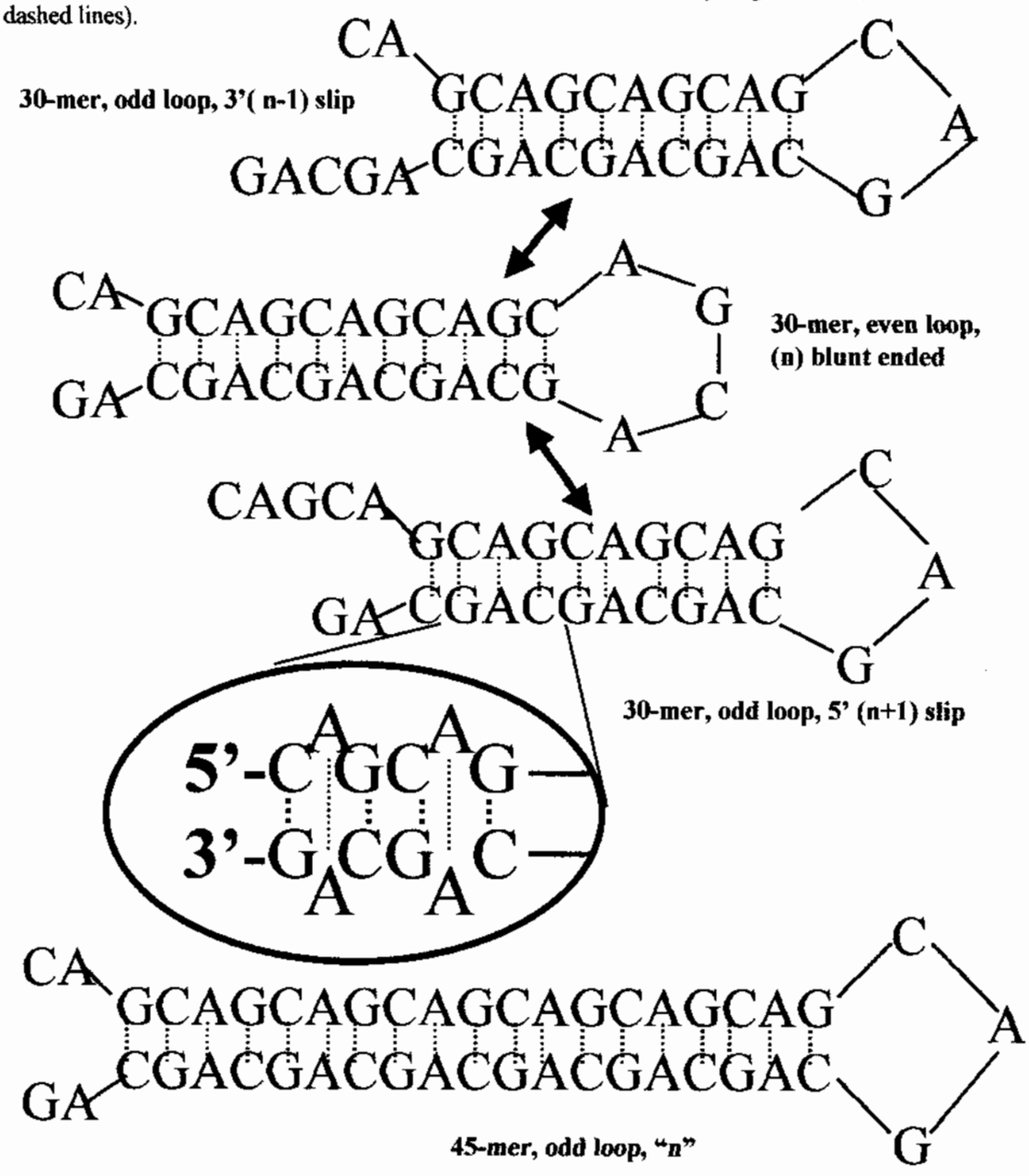
transitions that corresponded to the melting transitions of the quadruplex, then duplex structural forms. In all cases no multiple transitions were found at 115 mM Na^+.

To further define the type of structures formed, the oligomers were also examined with native gel electrophoresis. Native gel electrophoresis had shown, for all solitary sequences a single band with faster mobility than the corresponding complementary duplex (Figure 60, Figure 61). In no sample were there bands corresponding to a species with a molecular weight corresponding to the complementary duplex or higher. On the basis of native gel electrophoresis alone, the data indicate that all oligomers, from the 12-mer to the 75-mer across all (CNG)n sequences form unimolecular structures that are likely to be hairpins. As noted previously (4.2.6) the observed mobility of the solitary oligomers is higher than expected for a hairpin that is composed simply of n-mer/2 nucleotides, where n = oligomer length, but lower than that of the complementary duplex. There are several reasons for this apparently aberrant mobility. Native gel electrophoresis is a technique that separates biomolecules on the basis of both size and charge. Alterations to the size or charge affect the charge-to-size ratio and can therefore change the mobility of the molecular species in an electric field. These changes are dependent upon the nature of the change made to the molecule. All other variables remaining equal, if the charge of the molecule remains constant, then increasing the size of the molecule will decrease the mobility in a native electrophoretic gel. In contrast, a decrease in the apparent size will increase the mobility. If the hairpins formed by these sequences maintained full complementarity, then the expected size would approximate that of duplex DNA and the expected mobility would approximate that of and n/2 duplex

where n = the full length oligomer in nucleotides. However, as demanded by the nature of the triplet repeat sequence, the hairpins do not maintain full complementarity due to the mismatch located at every third nucleotide within the sequence. In addition, the loop region of the molecule also increases the apparent size by providing a large "bulge" at one end of the hairpin stem. Therefore, both the mismatches and the loop region have the net effect of increasing the overall size. The increase in size effectively reduces the mobility of the oligomer by reducing the charge to size ratio. Were these sequences forming mismatched duplexes, the amount of charge would be equivalent to that of a duplex, but the size would be larger due to the mismatches along the duplex length. The increase in apparent size of the molecule would decrease the mobility of the mismatched duplex relative to the complementary duplex, and produce a band that migrated with slower mobility than the complementary duplex. In all samples, none were found suggesting that the oligomers do not form higher order structures, or mismatched duplexes. Taken together, as a complete "picture," the experimental evidence in this work demonstrates that the (CNG)n oligomers, under the solution conditions specified - 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH 7.0, 100 mM NaCl , form hairpin structures with variable stem and loop regions (Figure 88).

Figure 88: The proposed hairpin structure of the CAG triplet repeat oligomer.

The proposed hairpin structures of the CAG10 and CAG15 triplet repeat oligomers are shown. Hydrogen bonds (3 per base-pair) are depicted by dashed lines, whereas sugar-phosphate bonds are depicted with solid lines. The CAG10 oligomer depicts the blunt-ended "n" structure, the 5' "n+1" slip structure, and the 3' "n-1" slip structure. For all structures shown below the minimal, but not necessarily the only, loop size is depicted. Also shown for detail is a concept drawing of the AA mismatch. The N.N mismatches will contain from one to two hydrogen bonds (red dashed lines).



As previously noted (4.2.1), the (CGG)n oligomers displayed unusual CD spectra that could not be readily assigned to a particular structural class of DNA. Consequently, determination of the (CGG)n oligomer structure relied upon electrophoretic, UV-melt, and concentration dependence studies. The (CGG)n oligomers did not form higher order structures, such as quadruplexes or slip structures within the limits of the electrophoretic measurement. All oligomers migrated as single bands on an electrophoretic gel under native conditions (Figure 61). In addition, no multiphasic transitions were observed by UV-thermal melting techniques (Figure 65). Finally, for CGG10 and CGG15, the slopes of the concentration-dependence curves were linear with near zero slopes (Figure 59, Table 17). With the CGG4 oligomer however, the trend of the concentration dependence plot was not clearly defined (Figure 59). In spite of this data set, the bulk of the experimental results are consistent with oligomers that form unimolecular hairpin structures. Based upon the secondary physical evidence (gel electrophoresis, concentration dependence) that suggests a hairpin structure, the unusual CD spectrum for the (CGG)n oligomers could be due to the orientation of the guanine-guanine mismatches within the stem region of the hairpin molecule. According to the solution NMR results of Zheng and coworkers, the mismatched guanine-guanine (G.G) bases can reorient about the anomeric carbon bond to enable hydrogen bonding between the mismatched nucleotides (Figure 89, Figure 90). In this configuration, the guanine on one strand is in the "*syn*" orientation (opposite that normally found in B-form DNA) whereas the guanine on the opposite strand is in the "*anti*" conformation. Reorientation about the glycosidic bond can alter the CD spectral signal and possibly introduce a small amount of negative ellipticity with the B-form DNA peak normally found within the 280 nm region of the

Figure 89: The structure of the adenine and cytosine mismatches.

The structures of the N.N adenine and cytosine mismatches are shown below. For both N.N mismatches the nucleotide bases are in the anti-conformation. Adenine makes one hydrogen bond with some base stacking interactions, whereas cytosine makes one hydrogen bond with limited base stacking interactions.

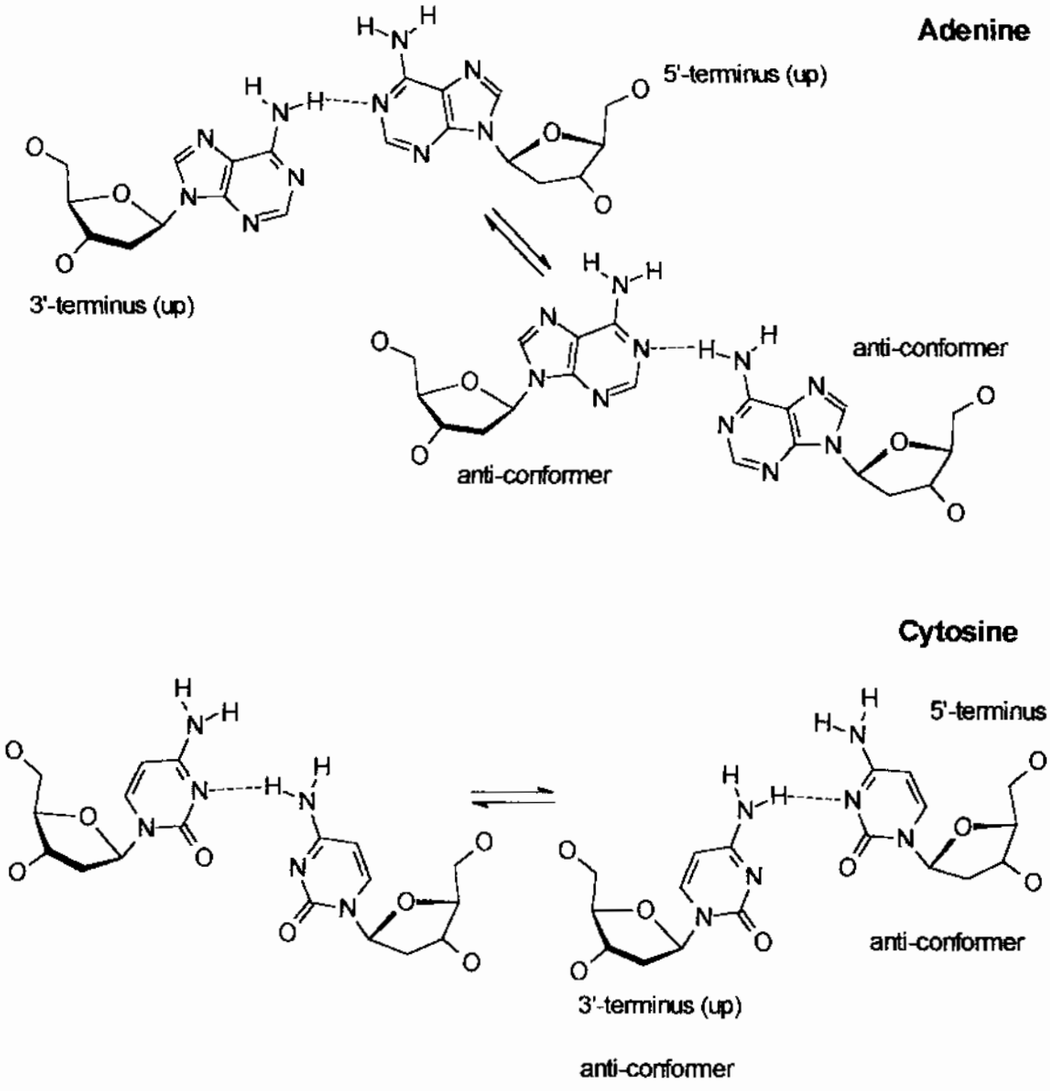
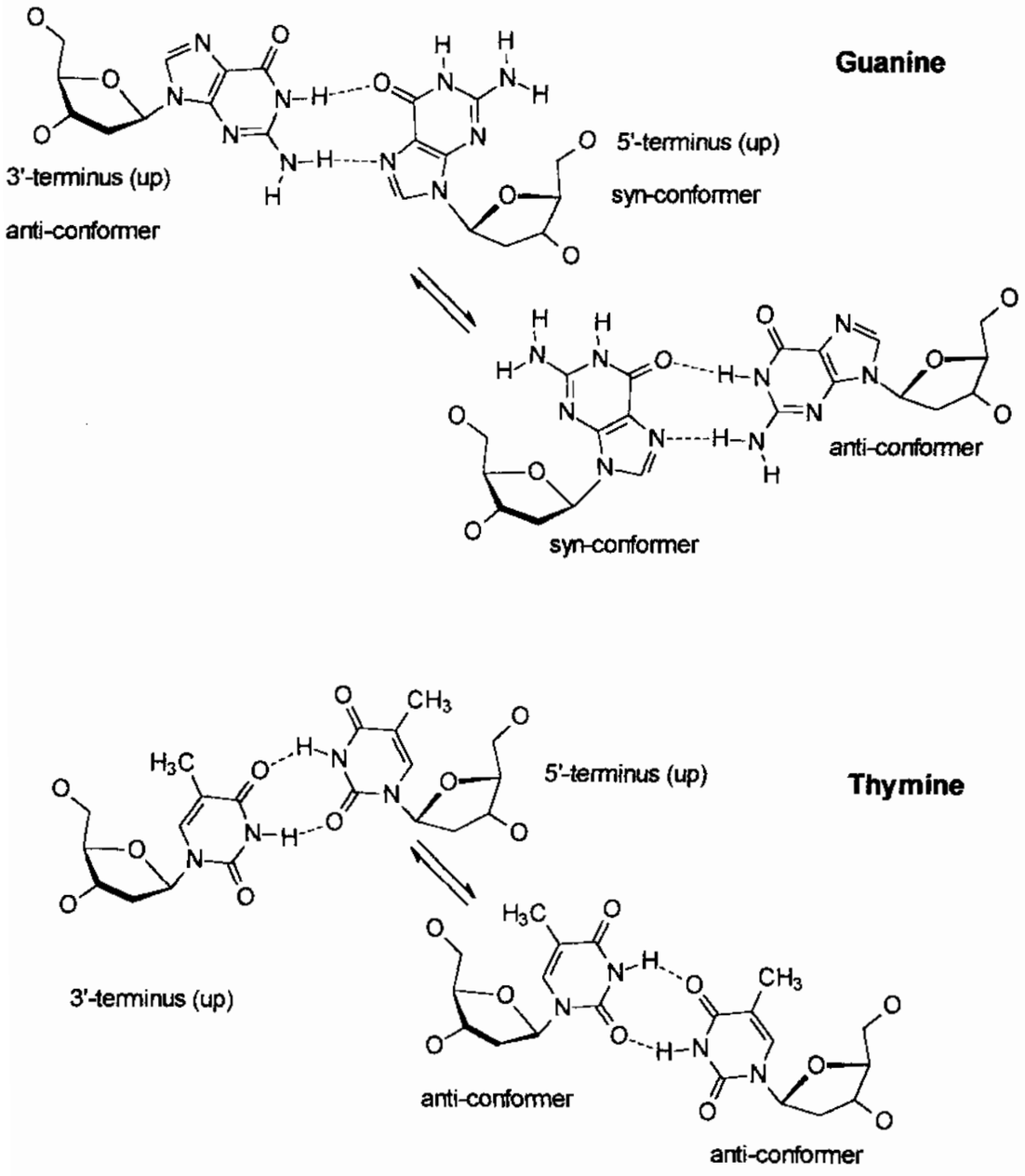


Figure 90 The structure of the guanine and thymine mismatches.

The structures of the N.N guanine and thymine mismatches are shown below. Guanine mismatches rapidly equilibrate between the anti- and syn conformations, can form two hydrogen bonds, and have extensive base-stacking interactions. Thymine "wobbles" about the plane of the nucleotide base,and can form two hydrogen bonds with some base-stacking interactions.



CD spectrum. The unusual CD spectra observed for the (CGG)n oligomers can, under this hypothesis, be explained as the superimposition of two types of CD spectra – the CD spectrum for the B-form DNA GC base pairs, and the CD spectrum for the *syn*-G.*anti*-G mismatch that produces a positive peak near 240 nm, and a negative peak (trough) near 280 nm (Figure 91). Several pieces of evidence from the CD spectral scans support this hypothesis. First, the only oligomer series with an unusual CD spectrum is also the only oligomer series containing a *syn.anti* mismatch. Second, the peak at 240 nm, not normally found in B-form DNA CD spectra and not found in the CD spectra of any other triplet repeat oligomer class other than (CGG)n, decreases in height as a function of temperature, and in relative proportion to the increase in height of the 280 nm peak. Third, the height of the peak occurring at approximately 280 nm in the 15 °C CD spectra of CGG15 appears to increase in height as a function of temperature. At 90 °C this peak resembles the CD spectra of all other oligomers at 90 °C (Figure 92). If the CD spectrum of the *syn*-guanine residue within the mismatch were to subtract from the peak normally found between 275 nm and 280 nm then the following properties would be expected. The height of the 275 nm should increase as a function of the denaturation state of the mismatch. Since the destruction of the "normal" CD spectral peak at 275 nm depends upon a purported interaction between the mismatch and the Watson-Crick base paired nucleotides both above and below the mismatch, then reduction of this interaction should return the CD spectral signal back to the type of spectra normally found for B-form DNA. At 15 °C the CD spectra within the 280 nm region does not resemble any known DNA form, but by 90 °C the 280 nm peak resembles that of single-stranded DNA at 90 °C. There is a net movement of the peak from negative to positive ellipticity because the G.G

Figure 91: A hypothetical deconvolution of the (CGG)n CD spectra

The CD spectra could be deconvoluted into the contributions of two different types of CD spectra. **A.** The CD spectrum for the Watson-Crick GC base pairs. **B.** The CD spectrum for the *syn*-G.*anti*-G non-Watson Crick hydrogen bonded pairs **C.** The observed CD spectrum composed of the sum of the spectra shown in A and B.

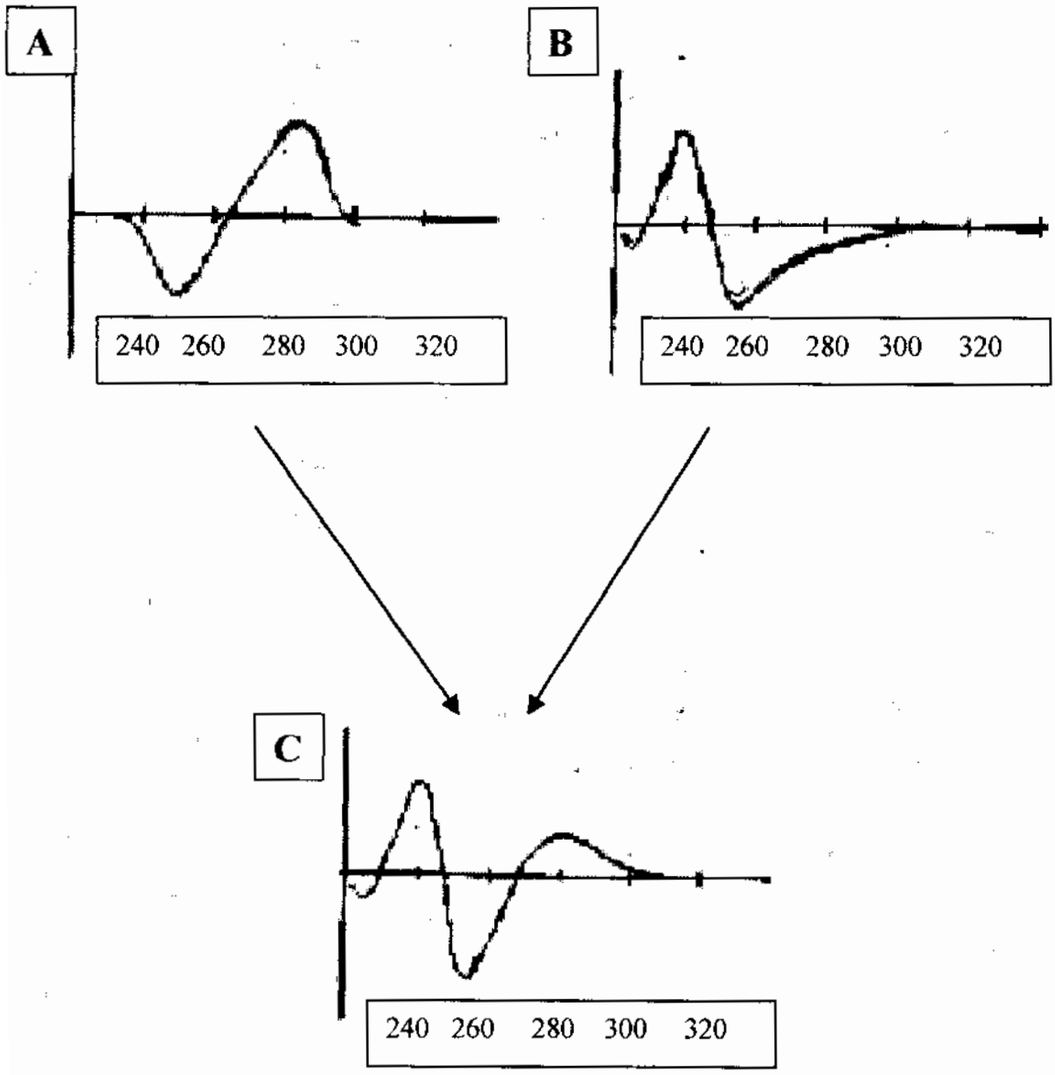
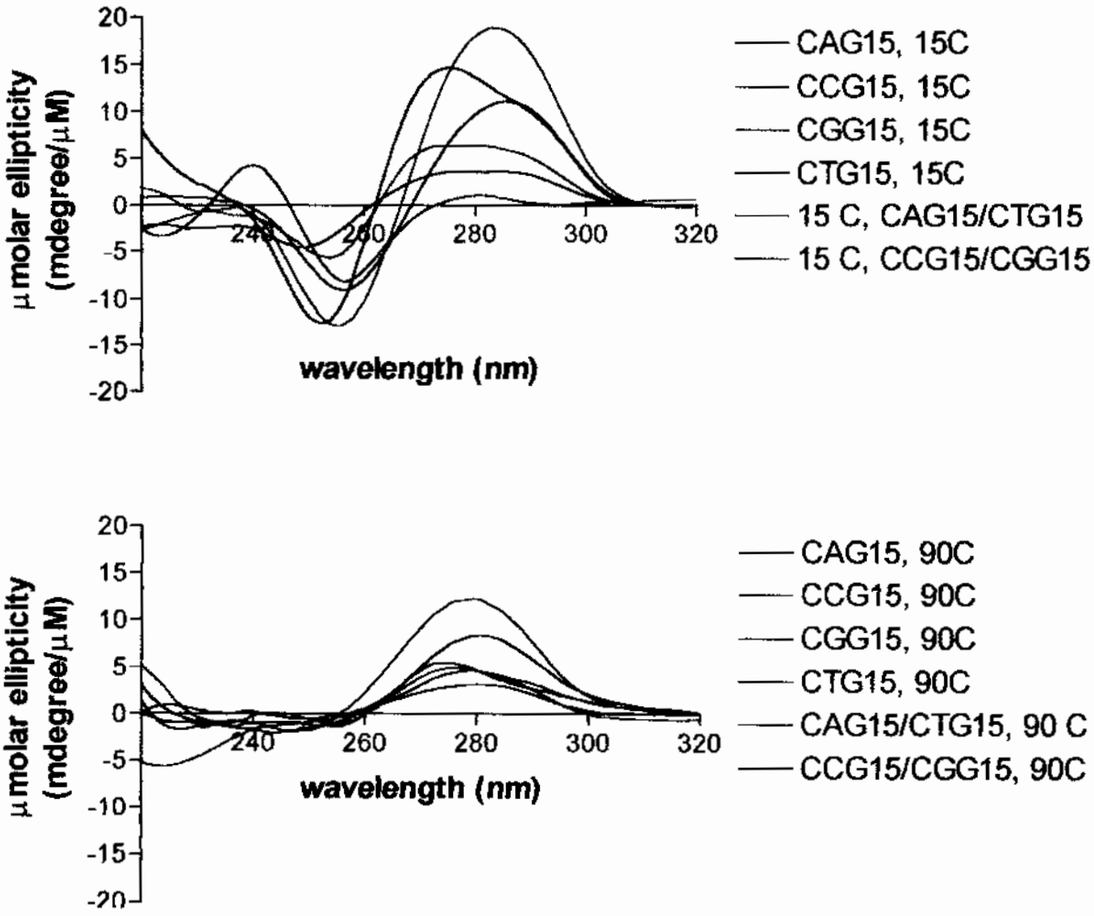


Figure 92: Comparison of CD spectra for structural comparison.

The 15 °C and 90 °C CD spectra of the solitary oligomers and duplexed oligomers in 5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH7.0, 100 mM NaCl are shown below for correlation with structural features. The 90 °C spectra serve as the reference point for comparison.



mismatch is thermodynamically less stable than the guanine-cytosine base pair. Denaturation of the DNA removed the subtractive influence of the mismatch and returns the CD peak to that observed for B-form DNA. In conclusion, the assignment of the (CGG)n oligomer structure can be made on the basis of native gel electrophoresis, DNA concentration dependence, and UV-melting techniques. The unusual CD spectra observed for the (CGG)n oligomers can possibly be the result of reorientation of the guanine nucleotides within the mismatch regions of the oligomer hairpin structure.

5.1.3 Complementary Oligomers form Duplexes

This dissertation shows that there are no unusual structural properties associated with complementary strands of triplet repeat DNA. On the basis of CD spectral data and native gel electrophoresis the complementary strands form B-form duplex DNA structures. The CD spectra for all lengths of the (CAG)n/(CTG)n and (CCG)n/(CGG)n duplexes, where n = 4, 10, 15, or 25 show typical B-from DNA spectra, with a peak centered near 275 - 280 nm, a cross over near 260 nm, and a trough at approximately 250 - 255 nm (Figure 42). The (CCG)n/(CGG)n CD spectra displayed a shoulder near 290 nm. In addition there was some variability as a function of length that suggested the sequences could undergo transitions with increasing length. Examination of the oligomers by native gel electrophoresis shows that complementary mixtures of (CAG)n/(CTG)n or (CCG)n/(CGG)n form structures that migrate with the expected size as the corresponding duplex DNA structure. In no sample were there observed changes in the mobility of (CCG)n/(CGG)n duplexes as a function of length. In contrast to the CD spectral data this indicates the (CCG)n/(CGG)n duplexes do not significantly change

size or charge, and imply little change in structure as the length increases. In addition, similar to the trends observed for the solitary oligomers, there were no bands observed in either sample set to indicate that higher order structures, such as quadruplexes were formed.

5.1.4 Proposed Structures

Thus far, the experimental results of this project could be summarized as follows. First, the CD spectral data indicates the solitary oligomers are forming structures from as short as 12 nucleotides to as long as 75 nucleotides. With the exception of the (CGG)n oligomers, most oligomers produced CD spectra similar to B-form DNA. In addition, the structure changes very little as a function of length. Second, the DNA concentration dependence experiments suggest that the DNA transition is unimolecular for most oligomers studied. Third, native gel electrophoresis results show that at 115 mM Na^+ the oligomers do not form structures that are consistent with either a mismatched duplex, or a higher order structure such as a quadruplex. Together, the results indicate that the solitary oligomers form unimolecular hairpins with a stem region similar to B-form DNA (Figure 88). Due to the alignment of the triplet repeat unit within such a structure it is impossible to achieve full complementarity between the "sense" and "antisense" legs of the hairpin. Every third position (nucleotides 5, 8, 11 etc) within the primary structure of the stem contains a mismatched nucleotide (Figure 88). The triplet repeat unit can be described as containing one mismatch sandwiched in between two complementary nucleotides. Because this structural feature is unlike B-form DNA, where all nucleotides within the sequence of the sense strand maintain full complementarity with the anti-sense

strand, this structural form is called B-like. While there are many properties in common between the B-form and B-like DNA, the location of a mismatch at every third nucleotide introduces a localized distortion or "bulge" to the backbone of the DNA helix. This distortion imparts to the hairpin molecule unique thermodynamic and structural properties in comparison to B-form DNA. As was previously shown, the hairpins are less stable than the fully base-paired duplex of similar length. Furthermore, Gellobolian and Bacolla have also shown that triplet repeat sequences containing mismatches possess greater flexibility than comparable duplexes of the same length.[30,31] In addition to a stem region containing mismatched nucleotides, the hairpins also contain a loop region opposite the 5' and 3' termini that enables formation of a self-complementary hairpin (Figure 88). The size of the loop is variable, and can range from a minimum of three nucleotides within the loop (the minimum number required to achieve a turn) to larger sizes that are dependent upon the overall stability of the hairpin structure. Loop structures within DNA molecules can have a destabilizing influence on the thermodynamic stability of the molecule due to torsional strain, in the case of small loops (n = 2 or 3), or exposure of hydrophobic groups in the case of large loops (n = 5 or more). Because the loop region contributes negatively to the thermodynamic stability of the DNA structure, the size of the loop will vary according to the prevailing thermodynamics of the hairpin. A balance must be struck between relaxation of loop strain and exposure of hydrophobic groups within the larger thermodynamic "picture." Hairpins with high stability, that can offset the destabilization of the loop region with a stabilizing contribution from the stem structure, will have comparably larger loops than those hairpin sequences with lower stabilities. The minimum size loop for an odd-numbered

repeat sequence in a blunt-ended alignment is three nucleotides, whereas for an even-numbered repeat sequence the minimum is four nucleotides. It is quite possible, however, that the average size for a loop region in an even-numbered DNA sequence is larger than three, due to N-N mismatch (where N = A, C, G, or T) located 5' to the GC base-pair closing the loop (Figure 88). In addition, it is also possible to change the size of the loop dependent upon the sequence alignment of the hairpin. Due to the sequence redundancy, there are multiple alignments possible for the hairpin structures proposed. In fact, the hairpin form is not the only structure that is theoretically possible with redundant sequences. Other forms, such as a cruciform, or hairpins with multiple mini-loops are also possible. However, the most likely structure formed will be the thermodynamically most stable structure that maximizes the hydrogen bonding between the sense and antisense legs of the hairpin. Such a structure is the blunt-ended hairpin. Slip-structures, where the alignment shifts over either in the 5' (N+1 slip) or 3' (N-1slip) direction by one repeat unit are also theoretically possible, but not thermodynamically favored beyond an n+1 or n-1 slip due to the low thermodynamic stability. It is quite possible, that in solution the oligomers are in rapid equilibrium between the N (blunt ended), N+1 (5' end slips in the 5' direction by one repeat unit), and N-1 (3' end slips by one repeat unit) structures, with the predominant form being the blunt ended alignment (Figure 88). On the basis of gel electrophoresis, the experimental results displayed no bands migrating below or above the main band corresponding to the solitary oligomer. Were N+1 or N-1 slip-structures formed in any significant concentration, then there would have appeared in the native gel a band appearing above the main oligomer band. The slip structure hairpin is slightly larger than the blunt-ended hairpin and should have a lower charge to size

ratio. Different structural forms imply different charge to size ratios due and should produce multiple bands in a native electrophoretic gel. For most oligomers, there was only a single band by native gel electrophoresis suggesting there is only one predominant form of the hairpin.

5.1.5 Differential Ion Binding Term Studies Show (CGG)n Most Compact

While the CD spectral studies, the native gel electrophoresis experiments, and the DNA concentration dependence measurements defined the gross details of the oligomer structures, the ancillary studies such as the salt effects and the temperature melting experiments, in combination with derivative studies, such as the CD signal/nucleotide defined the minor details of the structures formed by the oligomers examined. The differential ion binding term (DIBT) experiments produced data that were consistent with the thermodynamic stability trends determined by UV-melt techniques. This dissertation shows that the more stable triplet repeat sequences by UV-melt techniques also bound the most amount of sodium per oligomer. The magnitude of the differential ion binding term however, was unusually small for reasons that are not yet clear. Normally there are approximately 0.16 sodium ions released per phosphate upon melting duplex DNA. The observed range for all oligomers was between 0.12 for CAG4, and 1.54 for CGG10 (Table 18). These results were low by a factor of 3 - 5. Because of this experimental bias, analysis with this data set is restricted to trend analysis. Oligomers with high amounts of counterion condensation along the phosphate backbone of the molecule will release high levels of sodium upon melting. Because of the high amount of counterion condensation these oligomers will also have more compact structures due to reduction in the phosphate-phosphate repulsion. Thus, counterion condensation can be used as an

approximate measure of the relative compactness of two similar structures. This work shows, through the trend of the DIBT, that (CGG)n oligomers bound the most sodium ion per oligomer (Table 18). In terms of the amount of Na^+ ion released upon denaturation, the ion binding term progressed, from most to least, CGG15 > CTG15 > CAG15 > CCG15. These results indicate that part of the high stability observed for the (CGG)n oligomers is due to a reduction of the phosphate-phosphate charge repulsion along the DNA backbone. These results also suggest that the (CGG)n oligomers can assume a relatively compact structure when compared to the (CCG)n oligomers at the opposite end of the binding term ranking. The observed trends in the compactness of the oligomers, however, did not follow the expected solution CD spectral behavior as a function of increasing Na^+ concentration. Oligomers such as CAG15 displayed small incremental changes in the base stacking interactions with little dehydration as a function of sodium concentration. It was expected that a relatively open structure, as predicted by the DIBT would have more solvent exposed surfaces and more opportunity for the dehydration of the nucleotide bases. It appears, with respect to the (CAG)n oligomers, that changes occur to the backbone spacing, and these changes in turn reduce the base stacking interactions as a function of Na^+ concentration (Figure 55). The CCG15 oligomer behaved in manner consistent with a less compact structure as predicted by the DIBT. The oligomer under went both changes to the phosphate backbone, reflected by the loss of base stacking interactions, and the simultaneous hydration of the nucleotide bases, as reflected by the slight blue-shift in the peak maxima at approximately 285 nm (Figure 56). Interpretation of the CGG15 CD spectra as a function of salt is illustrative of the effects for the oligomer series. Because of the unusual CD spectra, for the (CGG)n

series, data interpretation of the CGG15 CD spectra followed the same kind of analysis previously employed. The CGG15 oligomer undergoes increased interactions, with respect to the 240 nm peak and increased destructive interference with respect to the trough at 260 nm and the peak at 280 nm. In all observed peaks there was little if any changes to the peak maxima or minima, suggesting very little changes to the hydration state of the nucleotide bases. These data suggest that the backbone of the (CGG)n oligomers retain the compact structures, and undergo changes that increase the overall base stacking interactions. The (CGG)n oligomers become more compact with increasing sodium concentration. The CD spectra for the CTG15 oligomers displayed much structural variation as a function of sodium concentration unlike the CD spectra observed for the CTG4 or CTG10 oligomers. This suggests that there is a length dependent change in solution behavior that is not observed with the other (CNG)n oligomers. The CTG15 oligomer undergoes significant changes to the sugar-phosphate backbone that realigns the nucleotide bases such that there is a net decrease in base stacking interactions with a simultaneous dehydration of the nucleotide. The shorter (CTG)n sequences show no such tendencies for diverse structural changes. Rather, these oligomers show a tendency for little changes in denaturation state or the hydration state of the molecule.

Overall, the DIBT and CD data set indirectly shows differences in the solvent accessibility of the nucleotide bases of triplet repeat sequences. Those oligomers with compact structures are also more thermodynamically stable, due to in part the reduction of phosphate-phosphate repulsion. Oligomers with compact and structures also have less

tendency for exposure of the hydrophobic nucleobases to the external solvent as shown by the relatively small amount of red-shift or blue-shift as a function of Na^+ concentration. In addition, because of this tendency, changes in structure occur by realignment the DNA backbone that causes an increase or decrease to specific base stacking interactions.

5.1.6 CD-melt Studies Show (CNG)n Structural Behavior

The CD-melt studies revealed a wealth of information about how the triplet repeat sequences melt from a structural perspective. For all data sets the CD spectra of the 45-mers were referenced because the CD spectra for this length was most representative of the CD spectra observed across all lengths of a triplet repeat class. The CD spectrum as a function of temperature reveals how an oligomer melts in response to temperature in a specific sense, or, in a general sense, to the changes undergone as the oligomers denature. A change in the nucleotide spacing (DNA rise) without loss of hydrogen bond interaction will change the CD spectrum by decreasing the base stacking interactions. The CD spectra should, as a function of denaturation, produce a series of peaks that become progressively lower. Similarly, loss of hydrogen bond interaction, produces not only a loss of base stacking interactions, but an increase in the hydration state of the DNA molecule due to an increase in the solvent accessibility of the nucleotide bases. The CD spectrum as a function of temperature for such a phenomena should show a decrease in base stacking interactions, with a concomitant shift, toward the blue region of the spectrum. As a function of temperature, the CD spectra of the (CAG)n oligomers showed little, if any, shift in the trough centered near 240 nm, while the peak centered near 275 nm displayed a slight blue-shift. The data suggests that the CAG15 oligomer, which is

representative of the trends observed for the shorter lengths examined, simultaneously denatures, through a loss of base stacking interactions, and hydrates as shown by the blue-shift of the peak at 270 nm. The amount of blue-shift observed is low and may be indicative of two different phenomena. First the (CAG) series of oligomers may have a more open, solvent accessible structure, such that denaturation when it does occur does not significantly changer the hydration state of an already highly hydrated molecule. The second possibility is the observed shift may be a property of the (CAG)n CD spectra, which could show little changes as a function of denaturation simply because adenine does not produce a strong CD signal. There were marked changes in the level of base-staking interactions with increasing temperature, suggesting that changes to the backbone occur first followed by loss of hydrogen bonding. As a function of temperature the CCG15 oligomer displayed a similar, but more exaggerated trend, with respect to the position of the peak near 285 nm and the trough near 255 nm (Figure 45). The positions of both the peak and trough incrementally shifted toward the blue region of the CD spectrum. Concurrent with the observed blue-shift the base stacking trends for the CCG15 oligomer were incrementally smaller for CCG15 than it was for the CAG15 oligomer, suggesting that the loss of interaction occurs slower for the CCG15 oligomer than for the CAG15 oligomer. This indicates that the CCG15 oligomer, and perhaps the (CCG)n oligomers as a whole should be more stable than the (CAG)n oligomers. In a general sense the data indicate that the (CCG)n oligomers undergo denaturation through a simultaneous loss of hydrogen bonding and base stacking interactions. The CD spectra of the CGG15 oligomer displayed little changes to the positions of the peak maxima and minima, until elevated temperatures were achieved (Figure 46). In addition, as the

temperature of the solution environment approached the melting transition of the CGG15 oligomer, there was a small but incremental decrease in the base stacking interactions, shown by the increase in the height of the peak near 280 nm and 240 nm. Overall, this was reflective of the overall stability of the molecule. For the CGG15 and (CGG)n oligomers in general, the CD spectral behavior is consistent with a high-stability oligomer that first undergoes changes to the rise of the molecule before loss of hydrogen bonding occurs. When the melting transition is approached, there is simultaneous loss of hydrogen bonding, and hydration of the newly exposed nucleotides. The CTG15 oligomer shows changes as a function of temperature that are similar to those observed for oligomer CGG15 (Figure 47) and are indicative of the thermodynamic stability of this triplet repeat class. The height of the peak at 285 nm and the depth of the trough at 255 nm remains relatively stable until 45 °C. In addition, the position of both the peak and the trough remain unchanged until 60 °C where the positions of the peak and trough undergo a blue-shift. Like the CGG15 oligomer, the data trends indicate that the CTG15 oligomer undergo very modest changes to the base stacking interactions, and hence undergo little if any change to the rise or twist of the molecule below the melting transition. Once the melting transition is approached, the oligomer undergoes a loss of hydrogen bonding, shown by the blue-shift, and a simultaneous loss of base stacking interactions.

5.2 Thermodynamics

5.2.1 Purpose

The length-dependence of triplet repeat DNA sequences with disease causation has been established by many laboratories within the field.[5,9,10] While some laboratories have studied the associated thermodynamic stability of triplet repeat DNA sequences,[22,29] none have undertaken a comprehensive and systematic study that examines both the length and sequence relationship. The goal for such a systematic thermodynamic study was similar to the structural studies – to define the database of thermodynamic properties and to determine if a correlation exists between stability and either the length or sequence context of triplet repeat DNA.

5.2.2 UV-Melts versus Calorimetry

Analysis of the UV-melt derived van't Hoff data versus the calorimetrically derived or calorimetrically derived van't Hoff data shows several significant differences between the data sets (Table 23, Table 26). First, as a function of length at 115 mM Na^+ the UV-melt enthalpic data are inconsistent with the magnitudes of the calorimetrically derived enthalpies. Second, the data trends of the UV-melt data are inconsistent within an oligomer series as a function of increasing length. On a theoretical basis it is expected that increasing the length of an oligomer will increase not only the enthalpy of interaction but the free energy of stability due to the additive effects of increased hydrogen bond formation and increased base stacking interactions. For the most of the data set, the UV-melt enthalpy trends show no such tendencies. While the (CAG)n oligomers for example displayed an increase in the enthalpy as function of length and are the only

oligomer series within the data set to do so, the (CCG)n and (CTG)n oligomers and the (CAG)n/(CTG)n duplexes show a net decrease in enthalpy as a function length, in spite of the fact that the melting transition temperature for all lengths studied increased with increasing length (Table 23). In addition, the enthalpy trend for the (CGG)n oligomers show a peak in the enthalpy data with increasing length.

Examination of the UV-melt data revealed systemic problems throughout the data set that could account for the erratic enthalpy data. In general, these problems were related to the manner through which the alpha slope of the melting transition was determined. Several of the samples, most notably the (CGG)n series had incomplete melting transitions, or transitions where the post-transition baseline was difficult to establish with the given software analysis package. Several other data sets had unusual transitions that could have been related to a multiple populations of states through out the melting measurement or multiple transitions of a single sample. Most problematic was determination of the linear range of the alpha plot that was the basis for calculation of the van't Hoff transition enthalpy plot. Small scale deviations in the slope of this line could produce different enthalpy values. Perhaps the most important reason for the failure of the UV-melt data set has to do with the underlying assumptions used to process the data. All UV-melt data assumes a two-state interaction where there are only two populations of oligomer – native or denatured. Deviations from the two-state interaction assumption will introduce a systematic error into the data, and more importantly, render useless the assumptions used for the determination of the equilibrium constant and van't Hoff enthalpy. In a non-two-state system the equilibrium position and concentrations between the native and

denatured populations is uncertain. Without accurate concentration data, the equilibrium constant, through which the van't Hoff enthalpy is calculated is uncertain. While the UV-melt data had served so well for other studies within this thesis, such as the salt dependence and differential ion binding term studies, this technique failed when comparisons across several different sample sets was attempted. The reason for this failure was the lack of a true two-state interaction between folded and unfolded DNA populations. Due to the redundant and self-complementary nature of triplet repeat sequences, it was likely that subpopulations of partially unfolded molecules existed. For this and many previously stated reasons the calorimetric data were used for all thermodynamic comparisons.

5.2.3 Thermodynamic Behavior of the Triplet Repeat Oligomers

5.2.3.1 Comparison of $\Delta H_{van't\ Hoff}$ vs. $\Delta H_{calorimetric}$ – Non-Two-State

Comparison of the calorimetrically derived van't Hoff data and the calorimetric data was attempted for two reasons : first, to determine the stateness of the transition, and second, to determine the size of the cooperative unit. Comparison between the calorimetric and van't Hoff data, derived from modeling studies of the calorimetric data, for all solitary oligomers revealed that the ΔHvh/ΔHcal ratios were significantly less than or greater than one. The data indicates that for most of oligomers examined, there were non-two-state interactions occurring as the oligomers melted from hairpin to duplex. Given the redundant nature of the sequences under consideration, it is quite possible for the DNA sequence to melt and reanneal to an open site on the same strand either downstream or upstream of the melting site. This would produce a population of multiple slip-structures

that do not melt as a single all-or-none population but rather as a distribution of several populations. In such a situation, the van't Hoff modeling of equilibrium-dependent measurements, such as UV-melt analysis, are not valid due to the uncertainty in the equilibrium concentrations of the reactants and products of the transition.

5.2.3.2 Sequence Context Comparisons

The triplet repeat oligomers displayed diverse thermodynamic behaviors. Comparison of the calorimetrically-derived free energy of the hairpin (HP) to single strand (SS) transition across the CNG10, CNG15 and CNG25 triplet repeat sequences shows a stability order, ranking from most to least stable, of CGG > CTG > CAG > CCG (Table 26). There are two deviations from this trend, occurring for the 12-mers and the 15-mers of the (CNG)n sequences. However, the trend for the 12-mers CGG4, CAG4, and CCG4 is consistent with the CNG15 and CNG25 trends after exclusion of the value for CTG4. With respect to the 15-mers, the opposite trend was observed, where the CAG5 oligomer had a higher stability than all other oligomers within the 15-mer series. Excluding the CAG5 oligomer as a data point shows that the stability trend observed is still consistent with that observed for the longer triplet repeat oligomers, namely CGG5 > CTG > CCG. The difference in the stability trends of the short-mers and the long-mers is due primarily to the nature of the integrated calorimetric peaks. Small changes to the integration limits of the small peaks, such as those observed for CAG4, could produce significantly different enthalpy values than similar changes to the integration limits of large peaks, such as those observed for CGG25. When deviant data points are removed from consideration, the observed trends are similar to those observed for the longer oligomers.

A comparison between the UV-melt derived stability order, in spite of the limitations of the UV-melt techniques, showed stability trends similar to that predicted by the calorimetric techniques. The UV-melt stability order for the 30-mers, 45-mers and 75-mers showed the general trend CGG > CTG > CAG > CCG. Both the calorimetric and UV-melt trends were similar to those found by Peyret who found that the order of mismatch stability for N.N mismatches at 37 °C to be G.G > T.T > A.A > C.C.[165]

5.2.3.3 Structural Basis for Observed Sequence Context Stability Profiles

Because of the sequence redundancy of the triplet repeat DNA, the stability of the molecules is predicted to be dependent on the thermodynamics and behavior of the N.N mismatches. Tied into this premise is the structure of the triplet repeat mismatches. The observed stability ranking, CGG > CTG > CAG > CTG, is consistent with what is known from a structural perspective for the triplet repeat sequences. Both Lane[166] and Zheng[15] studied the CGG triplet repeat sequence by solution NMR and found that the nucleotide base of the G.G mismatch can reorient about the glycosidic bond. The guanine base on the sense strand was in the *anti* conformation about the glycosidic bond, whereas the guanine base on the antisense strand was in the *syn* conformation. They also found rapid orientation exchange between opposite guanine mismatches. The reorientation of guanine within the G.G mismatch allows the formation of Hoogstein hydrogen-bonds between the N-1 of the *anti*-guanine and the O-6 of the *syn*-guanine and between the C-2 amino of the *anti*-guanine and the N-7 of the *syn*-guanine (Figure 90).[166] This reorientation allows two hydrogen bonds to form between the mismatches and adds to the overall stability of the DNA oligomer that otherwise would be destabilized by a G.G

mismatch. In addition to the stabilization achieved through hydrogen bonding, there is an additional stabilization derived from the burial of hydrophobic groups towards the core of the DNA molecule as well as base stacking interactions that can occur between the *anti*-, and to a lesser extent the *syn*-, guanine and the G·C base-pairs located immediately above and below the mismatch. The solution NMR of the T.T mismatches within the CTG triplet repeat class were studied by Arnold, who found that the mismatched thymines maintain an *anti*- orientation about the glycosidic bond.[167] However, on the basis of solution NMR they also found that the T.T mismatches are capable of forming a "wobble" base-pairs between O-4 of the "sense" strand and N-3 of the antisense strand and between N-5 of the sense strand and O-2 of the antisense strand (Figure 90). The wobble base pair allows horizontal shifting or wobble about plane of the nucleotide base to allow interaction between different hydrogen-bond donors and acceptors to occur. True to the nature of a wobble base-pair, Arnold observed that there is much interchange in the relative positions of the T.T mismatches. Like the G.G mismatch, the T.T mismatch confers additional stability to the oligomer, through the formation of two hydrogen bonds and base stacking interactions. However, unlike the G.G mismatches, there is less base stacking interaction to be gained from the relatively smaller thymine base. In addition, the variable nature of the thymine position confers upon the nucleotide a high degree of conformational entropy that would destabilize the oligonucleotide. Arnold also studied the A.A mismatches in the CAG sequence context and found that, like the T.T mismatches, a rapidly interchanging hydrogen-bonded structure between the bases of two adenine in the *anti* conformation occurred.[167] There was only one hydrogen bond proposed, that occurred between the C-4 amino of the sense strand, and the N-3 of

the antisense strand. Unlike the G.G and T.T mismatches, there was only one hydrogen bond formed, thus reducing, in a relative sense, the amount of stability gained. While it is likely that the A.A mismatches had more base stacking interactions than the T.T mismatches, overall the H-bonding was sufficiently reduced to lower the stability ranking of the A.A mismatches below that observed for the T.T mismatch. Finally, the C.C mismatch, studied by Arnold shows a propensity for wobble-pair formation, much like that observed for the T.T mismatch.[167] Like the T.T mismatch, both bases of the nucleotide are in the *anti-* conformation. There is hydrogen bond formation between N-3 of the sense strand and the C-4 amino of the antisense strand. Like all other wobble base pairs, there is much interchange in the relative position of the two adenine bases such that the identity of the hydrogen bond donor and acceptor changes depending upon the orientation of the wobble. Relative to the other mismatches within the stability series, the C.C mismatch confers the least amount of stability to the oligonucleotide for several reasons. First, the C.C mismatch can only form one hydrogen bond. Second, the small surface area of the cytosine base reduces the amount of base stacking interactions achievable between the mismatch and the Watson-Crick basepairs both above and below the mismatch. Third, there is an increase in the conformational entropy about the wobble base-pair that destabilizes the oligomer.

5.2.3.4 Length Comparisons

The calorimetrically-derived thermodynamic parameters were compared on the basis of length to determine the "macroscopic" (total length) and "microscopic" (nucleotide) trends in the thermodynamic parameters. To summarize, the following trends were

observed: first, small oligomers were stable enough to form hairpin structures; second, as the length of the oligomer increased the observed thermodynamic stability also increased; third, the stability (free energy), enthalpy and entropy all appeared to reach a plateau at approximately 45 nucleotides (Figure 76); fourth, on a length basis the enthalpy for the (CGG)n oligomers approached the theoretical limit of maximum hydrogen-bonding up to the 45-mer, whereas the (CAG)n oligomers were below the theoretical limit. On a nucleotide adjusted basis, the enthalpy/nucleotide was lower than that theoretically possible for the (CAG)n oligomers but for the (CGG)n oligomers it was close to the theoretical limit for all but the 75-mer; finally, the complementary duplex was more stable than the corresponding component oligomers in the hairpin state.

The first and second points are consistent with the basic properties of DNA that predict hydrogen bond interaction will increase as function of oligomer length. Single-stranded DNA is a thermodynamically unfavored state whereas the formation of a partially Watson-Crick complementary hairpin is a more favored hydrogen bonded state. With increased length of a redundant sequence there is a corresponding increase in the sites available for nucleation and hydrogen-bond interactions to occur. At the short end of the length spectrum, the (CGG)n oligomers approached the theoretical limit of enthalpy due to hydrogen bond interaction (Table 25, Table 26). On the surface this observation suggests that the CGG4 through CGG15 oligomers have achieved maximum hydrogen bond interactions. However, hydrogen bond interactions are not the sole component of the measured calorimetric enthalpy. Part of the measured calorimetric enthalpy is also due to base stacking interactions and charge stabilization that cannot be factored into a

theoretical enthalpy calculation because no context-relevant data sets exist. Therefore, while it appears that hydrogen-bond formation is maximized for the short to intermediate lengths of the (CGG)n oligomers, it is readily apparent that the theoretical enthalpy is higher and that the measured enthalpy does not reflect just hydrogen-bond formation. It is therefore likely, that while the (CGG)n oligomers possess the most hydrogen-bond and base stacking interactions of all oligomers examined, there is still some amount of unpaired mismatches or complementary nucleotides within the stem of the hairpin. At the opposite extreme of the thermodynamic spectrum are the (CCG)n oligomers, which have the lowest stability of all oligomers examined. Surprisingly the enthalpy for these oligomers is higher than that measured for the (CAG)n oligomers. Nevertheless, the enthalpy for this class of oligomers is far below the theoretical hydrogen bond limit for an oligomer of this class. This observation suggests that there is less base stacking and hydrogen bond interactions for the (CCG)n oligomers, and the (CAG)n oligomers as well. Comparison of the enthalpy trends between the (CGG)n and (CAG)n oligomers suggested that there is a difference in the extent of the hydrogen bond and base stacking interactions between the two classes of oligomers. The data also suggest that there is more structure associated with the (CGG)n oligomers, because of the high level of interaction than there is for the (CAG)n oligomers. Examination of the entropies for the hairpin to single-strand transition for the two data sets shows that the (CGG)n oligomers have higher entropy values (more negative entropy in going from single-strand to hairpin) than the (CAG)n oligomers (Table 26). The entropy for the hairpin to single strand transition gives a relative measure of the disorder of the ground state of the oligomer, and enables a ranking of the oligomers in terms of the disorder and/or flexibility of the

oligomer. The entropy of this process is the sum of the conformational flex of the molecule, the conformational freedom of the individual nucleotides, the localized melting/reannealing of the basepairs (a.k.a. "breathing" of the molecule) that includes the end-fraying of both the stem and loop regions of the hairpin. Assuming the single-stranded states are isoenergetic across all oligomers of the same length and contain the maximum allowable disorder, then oligomers with a high amount of disorder will have smaller entropy values in this direction relative to the oligomers with a low amount of disorder. Relative to the (CGG)n oligomers, the (CAG)n and the (CCG)n oligomers have a more disordered structure in the ground state at 37 °C. This observation is consistent with the prior data about these oligomers sets which on the basis of differential ion binding suggested (CGG)n bound more sodium and had a more compact structure, whereas (CAG)n bound less sodium and had a less compact structure.

Except for the (CGG)n oligomers, the enthalpy per nucleotide was lower for the short-length triplet repeat oligomers than for the intermediate-length triplet repeat oligomers ((Table 25 B, Table 29). After briefly approaching the theoretical limit, the enthalpy/nucleotide then diminished below that theoretically possible. Overall, except for the (CGG)n oligomers, none of the triplet repeat sequences examined approached the theoretical enthalpy limit due to hydrogen bonding. If the total enthalpy were the sum of the individual enthalpies for each triplet repeat unit, and assuming the structure maximizes hydrogen–bonding and base stacking interactions, then the enthalpy per nucleotide should remain constant, within experimental error, as function of length. The additive effects of hydrogen-bond and base stacking interactions become averaged across

the entire oligomer as the length increases. The fact that most oligomer sequences are below that theoretically possible suggests, that the (CNG)n sequences are biased towards instability.

When the stability of the complementary duplexes was compared to the stability of the component solitary oligomers an interesting trend was found (Figure 76). The free energy of the component hairpin oligomers was less favorable than the free energy of the duplex. This trend suggests, as expected, that the nucleotide bases of the duplex can form more hydrogen-bonds and participate in greater base stacking interactions than the component hairpin oligomers (Summary, Figure 90).

5.2.4 The Stability Plateau and Triplet Repeat DNA Expansion

This work shows, for the first time, the relationship between length and stability. Examination of the length relationship of the stability, enthalpy, and entropy for the hairpin to single strand transition revealed a plateau in the data trend beginning at approximately 45 nucleotides. In addition, while the stability of the component oligomers tended to plateau the duplex oligomers displayed an additive behavior where the stability increased in a near-linear fashion over the range of lengths examined. For a triplet repeat DNA sequence to expand, there has to be a metastable state that is intermediate in stability between the Watson-Crick duplex DNA at one extreme and the single-strand state at the opposite extreme. Stable hairpin formation requires that the parent DNA strand, once separated into the component single-strands, can readily form a hairpin structure to prevent reannealing between the two complementary strands from

5.3 Kinetics

5.3.1 Purpose

Gacy and McMurray were the first to report a kinetic difference in the reassociation rates between short and long triplet repeat DNA sequences.[29] They further hypothesized that the difference could account for triplet repeat DNA expansion. We examined the kinetic relationship between all (CNG)n classes of triplet repeat DNA as a function of length to determine if the relationship reported by Gacy and McMurry was consistent across all sequence types and lengths. In addition, we sought to extend their work by determining if there exists a relationship between the energies of activation for the hairpin reassociation process. In addition, we also sought to define the nature of the transition state for these oligomers.

5.3.2 Kinetics of the (CNG)n Triplet Repeat Oligomers

The triplet repeat oligomer hairpins once formed will eventually reanneal to a complementary duplex because the duplex state is more favored in a thermodynamic sense than the hairpin state. Gacy found a kinetic difference in the observed rate constant for the hairpin to duplex reannealing process suggesting that there is a mechanistic difference between the reannealing mechanisms of the triplet repeat oligomers.[29] The rate constants of all (CNG)n triplet oligomers, where n = 4, 10, 15, and 25 were measured as a function of temperature to determine the energy of activation for the hairpin to duplex reannealing process. The results show that the CAG/CTG hairpins reanneal with slower kinetics and a higher energy of activation than the CCG/CGG hairpins (Figure 82). When the data were replotted as an Eyring plot to determine the transition state

parameters, similar trends to those observed for the energy of activation were found. The transition enthalpy was lower for the CCG/CGG pair than it was for the CAG/CTG pair. On a length adjusted basis, the short-length oligomers reannealed with faster observed rate constants than did the long-length oligomers(Figure 83). When further mechanistic information was sought, the rate constants of the (CNG)n triplet repeat oligomers, where n = 4, 10, or 15 were measured as a function of sodium concentration between 15 mM and 500 mM Na^+ using the theoretical framework of Azivonas and Kearns.[38] The results show that the short-length oligomers reanneal with faster kinetics as a function of sodium concentration than do the long-length oligomers (Figure 85). In addition, energy of activation for the reassociation of the CAG/CTG pair exceeded the summed hairpin stability of both components. The energy of activation for the CCG/CGG reassociation however exceeded the summed stability for only the 12-mer hairpin pair (Figure 86).

5.3.3 The Transition State - E_a, ΔH, ΔS

5.3.3.1 The Nature of the Transition State

For most DNA reannealing processes, the rate-limiting step is nucleation or the process through which a DNA strand undergoes hydrogen-bond interactions with a complementary strand. To initiate hydrogen-bond formation the incoming nucleotide must overcome the steric and electrostatic repulsion forces, as well as hydration between the nucleotides in addition to achieving an overall stable thermodynamic state. Once nucleated, the reannealing process assumes kinetics that are faster than the nucleation process, because of cooperativity in the DNA – once properly oriented subsequent annealing becomes easier for the downstream nucleotides. In addition, the additive

effects of thermodynamic stability, that propagates down the DNA chain as the more nucleotides anneal, becomes larger helping to stabilize the structure and reduce the probability of strand dissociation due to end-fraying. In a similar fashion, the transition state of the triplet repeat HP to DUP association process is defined as the nucleation step between two complementary hairpin oligomers. The most probable nucleation-site for the reannealing process is the 5' and 3' termini of the hairpin stem. The 5' and 3' termini of the hairpin is the site of much end-fraying due to instability arising from conformational freedom near the ends of the DNA molecule. It has been shown that triplet repeat sequence DNA with the mismatch located at every third nucleotide is more unstable than duplex DNA and as a consequence it is expected to have more end-fraying than duplex DNA. The energy of activation, and more importantly the Eyring plot data describe the transition state through which the frayed-ends of the oligomer pairs anneal and form duplex DNA.

The transition state of the hairpin to duplex association process was characterized by studying the energy of activation, Ea, and the transition state enthalpy, ΔH_{ts} , and entropy, ΔS_{ts}, for the hairpin to duplex reassociation process. With respect to sequence context the energy of activation for the (CAG)n/(CTG)n series of oligomers was lower than that observed for the (CCG)n/(CGG)n series of oligomers by approximately 10-15 kcal/mol. This result indicates that the (CAG)n/(CTG)n oligomers reanneal with relatively slower kinetics than (CCG)n/(CGG)n oligomers (Figure 82, Table 33). This also indicated that there were underlying thermodynamic differences in the manner in which the two sequence pairs approached the transition state that could possibly be related to the DNA

sequence context. When the $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ were compared trends consistent with the energy of activation studies were found. The enthalpy of activation was higher for the (CAG)n/(CTG)n oligomer pairs than for the (CCG)n/(CGG)n oligomer pairs by approximately 10-15 kcal/mol. In similar fashion, the entropy of activation was higher for the (CAG)n/(CTG)n oligomers than for the (CCG)n/(CGG)n oligomers by approximately 20-30 cal/molK. Both experimental measurements indicate that there is more bond breaking and more disorder as the (CAG)n/(CTG)n oligomer pair approaches the transition state than for the (CCG)n/(CGG)n oligomer pairs.

From the start of this process the reactants have different ground states. On the basis of structural and thermodynamic studies it has been shown that CGG is highly stable and ordered, whereas the complement CCG is highly disordered. CTG on the other hand is more ordered and stable than CAG, but not more so than CGG. The oligomers that are the most stable undergo the least amount of end-fraying, whereas the least stable oligomers (CAG,CTG) undergo the most amount of end fraying. As these oligomers undergo end fraying, there is an energetic price to pay for all of the instability. The nucleotides with the most end fraying or conformational disorder associated with the molecule cannot find the correct orientation for nucleation making reannealing difficult and as a consequence have a slower association rates.

When examined as a function of length the kinetic trends were consistent with what is known about the thermodynamic behavior for these sequences were observed. For the short oligomer pairs the energy of activation was lower, the activation enthalpy was

lower, and the entropy of activation indicated that the transition state was more ordered than these same parameters measured for the long oligomers. This observation is consistent with an oligomer that becomes more disordered, or has more conformational freedom as a function of length. As the short oligomer approach the transition state, there is considerably less bond breakage occurring due primarily to the fact that here are less bonds to break in the shorter sequences to achieve the transition state. Long oligomers on the other hand have larger energy value due to the fact that there are more bonds available for scission and there is more conformational freedom associated with the molecule.

5.3.3.2 Salt Effects on the Observed Rate Constants

To gain additional mechanistic details about the nature of the transition state the effects of sodium ion concentration were measured. According to the theoretical framework of Azivonas and Kearns, the reassociation two hairpins forms of DNA can occur through a single-strand or cruciform intermediates.[38] If the duplex association process occurred through a transition state that resembled a single-stranded intermediate, then the observed rate constants would be positively affected by an increase in Na^+ concentration. Single-stranded DNA is destabilized by increasing Na^+ concentration, which accelerates the decomposition of the single-strand transition state into the more thermodynamically stable duplex form. On the other hand, if the duplex association process occurred through a cruciform intermediate, then the duplex reassociation process would either be unaffected or inhibited by increasing Na^+ concentration. The intermediate resembles duplex DNA that is stabilized by Na^+. The decay of the intermediate can proceed in two directions, either formation of the duplex DNA, or decay back into the hairpin forms.

Decay back into the hairpin form accelerates the reverse reaction, and depending upon the magnitude of this process will either show an inhibition of the forward reaction or no effect on the observed reaction rate.

The results show that the short-length oligomers reanneal with faster kinetics as a function of sodium concentration than do the long-length oligomers (Figure 85). This indicates that short-length oligomers undergo a transition state that is more single-strand like, where as long-length oligomers under go a transition state that is more cruciform like.

As an additional test of the mechanistic approach of the two oligomer pairs, the measured energy of activation for the hairpin to duplex association process at 115 mM Na^+ was compared to the thermodynamic stability of the component hairpin oligomers. If the duplex association "reaction" proceeded through a single-strand intermediate, then the energy of activation would exceed the summed thermodynamic stability of the hairpin pair due to the large number of hydrogen bonds that are simultaneously being broken.[38] On the other hand, if the hairpin to duplex association process proceeded through a cruciform intermediate then the energy of activation would be below the summed thermodynamic stability of the two hairpin strands due to the breakage of only a few hydrogen bonds at a time.[38] When the kinetics data set was examined as a function of strand sequence context, it was observed that the energy of activation for the reassociation of the CAG/CTG pair exceeded the summed hairpin stability of both

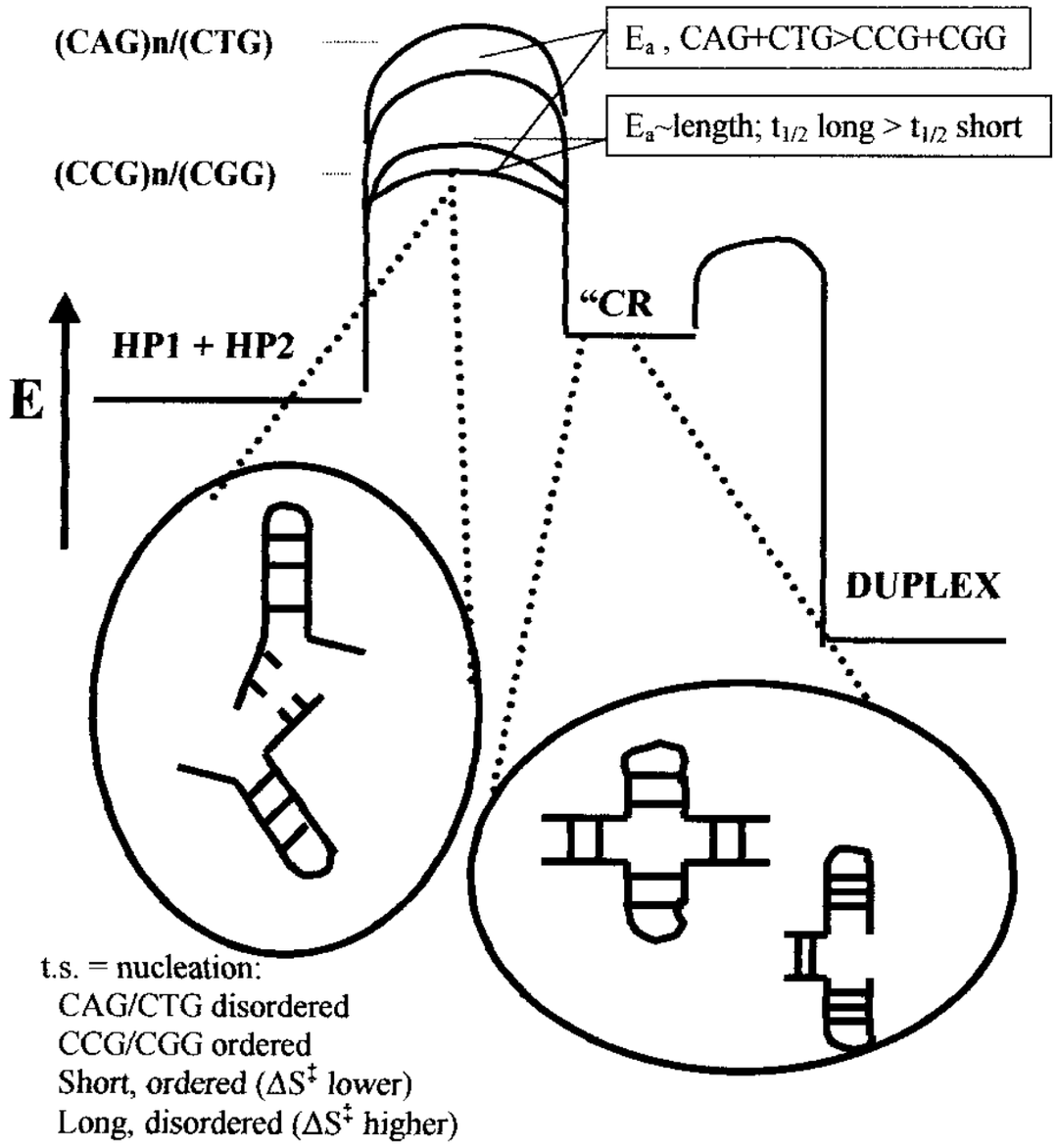
components. The energy of activation for the CCG/CGG reassociation, however, exceeded the summed stability for only the 12-mer hairpin pair (Figure 86). This result indicates that the (CAG)n/(CTG)n hairpin pairs associate to form a duplex through a transition state that is more single-stranded than the (CCG)n/(CGG)n hairpin pairs. When examined on the basis of length (Figure 86) the data trend shows that (CAG)n/(CTG)n the thermodynamic stability of the (CAG)n/(CTG)n oligomers does not exceed the energy of activation until approximately 100 nucleotides whereas the thermodynamic stability for the (CCG)n/(CGG)n surpasses the energy of activation at approximately 25 nucleotides. Consistent with the transition enthalpy and entropy data, these results indicate that the (CCG)n/(CGG)n oligomer pairs proceed through a more ordered, or cruciform intermediate, whereas the (CAG)n/(CTG)n oligomer pairs proceed through a disordered or single-strand like intermediate (Figure 93).

5.3.4 Some Mechanistic Details of the Transition State

5.3.4.1 How Each Oligomer Approaches The Transition State – E_a, ΔH, ΔS

As a whole, the results indicate that the two oligomer pairs proceed through slightly different association pathways that have down stream implications for disease causation. The entropy of activation for the (CAG)n/(CTG)n oligomer pairs indicate that the transition state is more disordered than the (CCG)n/(CGG)n transition state. This observation is consistent with the Na^+ concentration studies that indicate the transition state for the (CAG)n/(CTG)n oligomer pairs are more single-stranded in nature whereas for the (CCG)n/(CGG)n oligomer pairs it is more cruciform-like in nature. Similarly,

Figure 93: An energetics view of the kinetic mechanism.

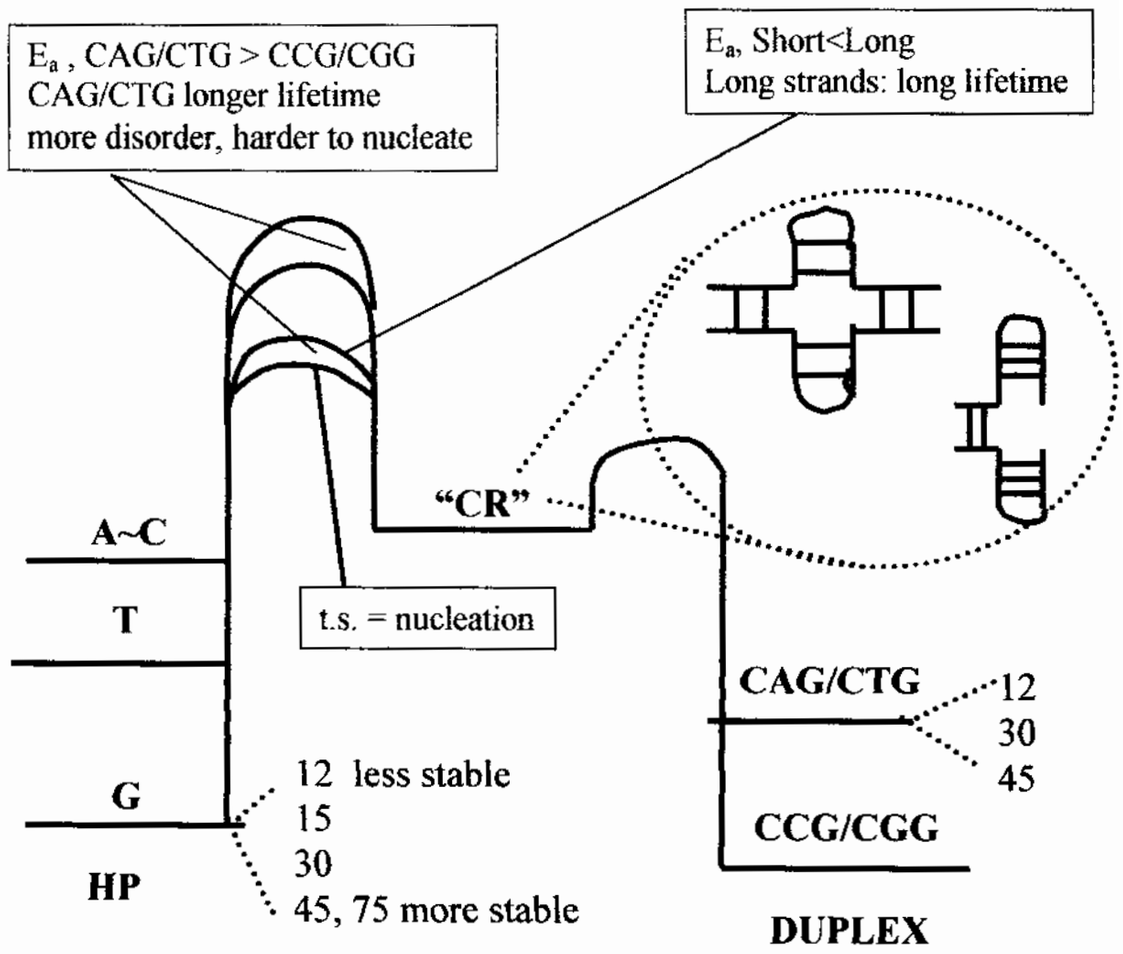


comparison of the enthalpy of activation shows that the (CAG)n/(CTG)n oligomers have more bond breakage than the (CCG)n/(CGG)n oligomer pairs in the transition state. This result is also consistent with the Na^+ concentration studies that indicate more bond breakage is occurring in the transition state of the (CAG)n/(CTG)n than for the (CCG)n/(CGG)n oligomer pairs. When examined on the basis of length, both the enthalpy and entropy of activation is higher for the long length oligomer strands than it is for the shorter strands suggesting that the transition states for the short strands are more ordered than the transition states for the longer oligomers. This result is, on initial consideration, inconsistent with the Na^+ effect studies that indicate the transition states of the short strands are more single-stranded in nature than the transition state of longer strands. The apparent contradiction in the enthalpy and entropy terms could possibly be due to the parameter measured in each technique. The entropy of activation views the effects of the entire molecule – larger oligomers have more entropy in the transition state due to the fact that a larger size brings with it more conformational freedom and more entropy than smaller oligomers. So while the transition state for the longer oligomer pairs may resemble a cruciform intermediate, the entropy is larger due to the larger size of the transition state under consideration.

In general terms, the transition states for the (CAG)n/(CTG)n and (CCG)n/(CGG)n pairs are dependent upon the sequence identity and length of the oligomer pair. It is known for duplex DNA sequences in solution that the 5' and 3' ends of the structure are unstable and will undergo a process referred to as end-fraying. End-fraying is a process of localized DNA melting at the 5' and 3' termini, followed by denaturation up to

approximately two or three basepairs. Duplex DNA in solution is therefore a mixed structure, containing a duplex DNA region and a very small portion near the ends that are in equilibrium with a single-stranded state. The single-stranded DNA ends, however, are not a stable state and eventually the free single-stranded ends will nucleate forming the fully formed duplex. End-fraying is the result of thermodynamic instability of the 5' and 3' nucleotides, due in part to the thermal fluctuations of the oligomer and is a rapid equilibrium process. It is proposed, on the basis of what is known about the properties of DNA that the hairpin oligomers undergo extensive end-fraying due to the thermodynamic instability of the N.N mismatch. In addition, the (CAG)n/(CTG)n oligomers pairs, on the basis of the data presented undergo more end-fraying than the (CCG)n/(CGG)n oligomers. The "free" single-stranded ends can then find an intramolecular nucleation partner and decay back to a hairpin or an intermolecular partner, to eventually form a duplex DNA molecule. In the case of the latter, the oligomers form either an intermediate that is similar to a single-stranded species, in the case of the (CAG)n/(CTG)n oligomers or, in the case of (CCG)n/(CGG)n oligomers, form an intermediate that is cruciform-like. Because the (CAG)n/(CTG)n oligomer pairs undergo an intermediate that is more single-stranded in nature, based upon entropy of activation data and Na^+ effect data, the likelihood for nucleation is lower and the observed rate constants slower. On the other hand, the (CCG)n/(CGG)n oligomer pairs have a more ordered transition state, allowing the nucleation process to occur rapidly (Figure 94).

Figure 94: A view of the proposed pathway for each oligomer pair.



5.4 Summary and the Biophysical Basis

5.4.1 Expansion of Triplet Repeat DNA Sequences – A Biophysical View

At the cellular level, the expansion of triplet DNA begins during replication. Replication is a complex process that involves a multitude of DNA processing enzymes, DNA synthesis proteins, and other proteins that maintain the single stranded nature of the unwound DNA. Throughout this process the parent DNA strand is first unwound and separated into single strands through the combined actions of DNA helicase and other processing enzymes collectively referred to as the DNA_{pol} complex. It is during the unwinding process that the expansion of triplet repeat DNA begins. According to Gellobolian the unwinding process introduces an extraordinary amount of superhelical torsion about the unwound DNA, such that the DNA_{pol} complex can no longer process the parent DNA strand (Figure 6). The DNA_{pol} complex then dissociates from the parent DNA and allows, through a process that has not been well-characterized, the formation of self-complementary structures on the daughter DNA strands. The redundant nature of the triplet repeat sequences allows a hairpin formation process that is faster than the binding kinetics of the single-strand binding proteins, whose function is to maintain any single-strand region of DNA as an uncomplemented single-strand. The structures formed during replication have been characterized by others as hairpins.[10,16]

This work shows, on the basis of circular dichroism, native gel electrophoresis and DNA concentration studies of the (CNG)n triplet repeat oligomers (where N= A,C, G, or T, and n = 4, 10, 15, or 25) that triplet repeat sequences readily form hairpin structures for all

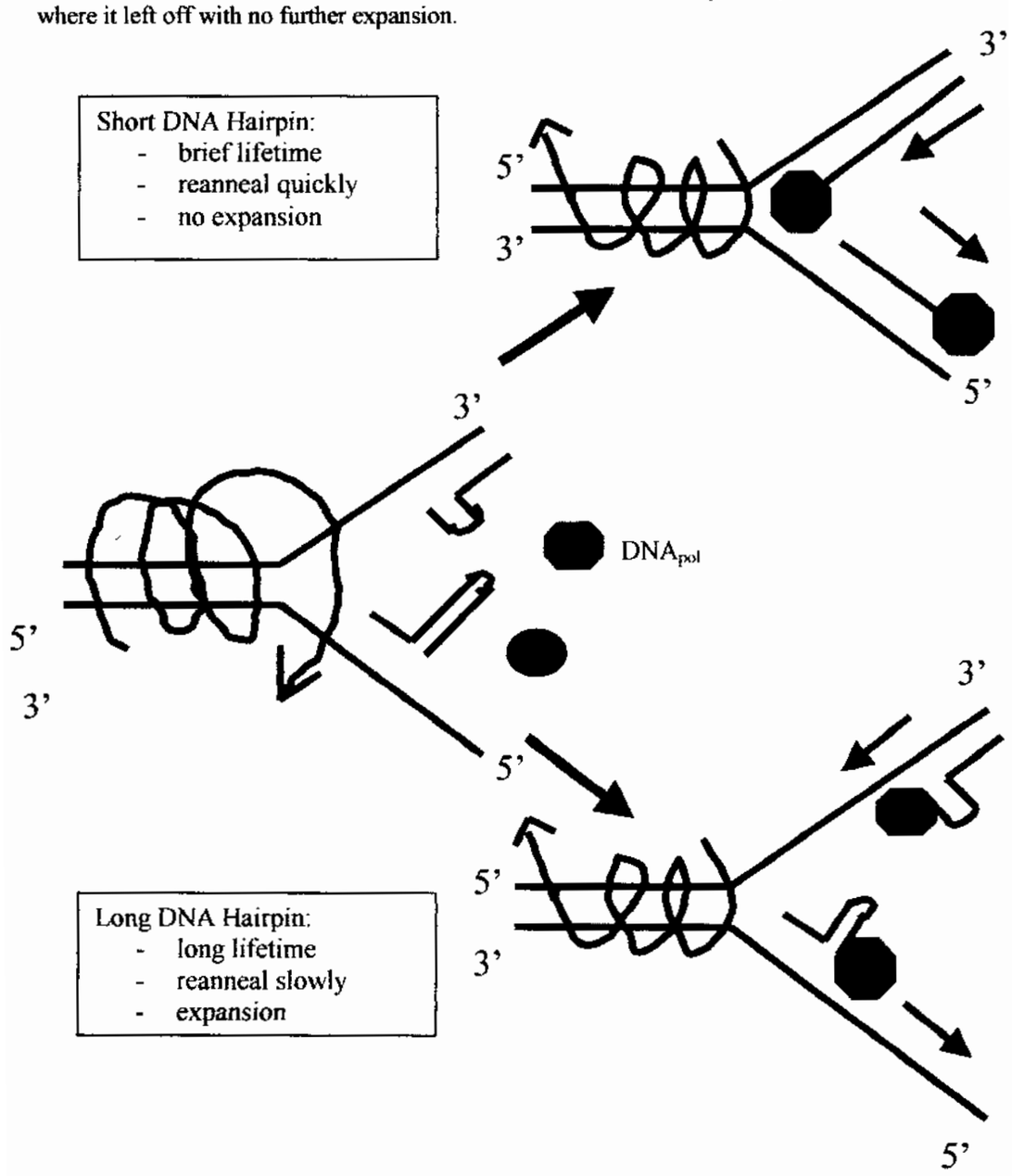
triplet repeat classes and lengths studied. These structures readily form *in vitro* under solution conditions (5 mM NaH_2PO_4, 5 mM Na_2HPO_4, 0.1 mM EDTA, pH7.0, 100 mM NaCl, 37 °C) that are similar to the physiological conditions. It is therefore concluded, with the assumption that there is direct correlation between what is observed *in vitro* with the oligomer model system and the behavior of these sequences *in vivo*, that the triplet repeat DNA sequences can readily form hairpin structures within the nucleus during replication. It is surprising that oligomers as short as 12 nucleotides long can form relatively stable DNA hairpins *in vitro*. Whether or not this is the case *in vivo*, within the cell nucleus is uncertain given the high probability that the hairpin end will find a nucleation partner on the parent strand and then realign. Hairpin formation is accompanied by a positional shift of the 3'-terminus of the daughter DNA strand such that a region of complementary parental DNA that once base-paired with the daughter strand is now exposed. Once formed the hairpin structures have two possible fates that dictate the downstream disease pathogenesis. These fates are directly dependent upon the timing of the ensuing processes. As mentioned previously, (2.5.5) the superhelical torsion leads to dissociation of the DNA_{pol} complex from the parent DNA strand which allows the DNA to release the tension. During this release there is hairpin formation, due perhaps to the high amount of energy imparted to the strands during the unwinding process, or perhaps due to the fact that there is no longer a polymerase associated with the strands. Subsequent to release of the superhelical torsion and hairpin formation, the DNA_{pol} complex will reassociate to the parent DNA strand. The location of the DNA_{pol} complex reassociation will be dependent upon both the thermodynamic stability and the reassociation kinetics of the hairpins. If the hairpin remains as a quasi-stable species,

then the DNA_{pol} complex reassociates at a location that is downstream with respect to the parental strand where the complex dissociated. Upon resumption of DNA synthesis, the DNA_{pol} complex reiteratively synthesizes a previously templated region of DNA, which leads to an expansion in the triplet repeat copy number (Figure 95). If end-fraying of the hairpin occurs and the free 3'-terminus of the hairpin nucleates with the parent DNA strand, then it is probable that the daughter strand will realign and form a duplex DNA or other structure closely resembling duplex DNA. In such a situation the reassociation of the DNA_{pol} complex occurs at a location at or near where the complex dissociated. Upon resumption of synthesis, there is no expansion in the triplet repeat sequence copy number.

On the basis of the data presented within this dissertation, the (CNG)n triplet repeat DNA sequences present a multifaceted dependence on the thermodynamic and kinetic properties that predict whether or not a particular sequence will lead to a stable hairpin that will lead further expansion *in vivo*. This work shows, from a thermodynamic perspective, calorimetric data suggesting that the stabilities of the hairpins are both length and sequence-context dependent. It is not surprising to find a sequence context dependency for the hairpin triplet repeat sequences considering that duplex DNA also demonstrates a sequence context and that the level of interaction in the N.N mismatch differs depending upon the sequence. Sequences with G.G mismatches within the stem region of the hairpin also had the highest stability of all oligomer sequences due to the formation of two hydrogen bonds and extensive base stacking interactions (Figure 90). Sequences with C.C mismatches on the other hand had the least stability due to the formation of only one hydrogen bond in the mismatch and limited base stacking

Figure 95: Biological consequences of the renaturation properties.

The biological consequences of the biophysical properties of triplet repeat DNA sequences. Sequences with small hairpins are less likely to form stable structures, and will reanneal back to the complementary partner faster than long stretches of triplet repeat DNA. Long sequences are more likely to form stable structures and will reanneal to the complementary duplex slower than short sequences. However, this stability is length dependent – beyond 45 nucleotides, the thermodynamic stability decreases, making increasing the probability for additional loop or bulge formation. If the structure remains prior to reassociation of the DNA_{pol} complex, then expansion will occur. If however, the structure does not remain, then the DNA_{pol} complex reassociates where it left off with no further expansion.



interactions. Considering the similarity between the (CNG) sequences - a 5'-CG-3' on the sense strand base paired to a 5'-CG-3' on antisense strand - it is the thermodynamic properties of the N.N mismatches that in large part drive the stability of the hairpin structures.

On the basis of thermodynamics alone, the CAG and in particular the CCG sequences were shown *in vitro* to have the lowest stability and are therefore predicted to have the least stable *in vivo* structures. If structure stability correlates positively with expansion probability then the CAG and CCG triplet repeat sequences would be predicted to be expanded the least and the CTG and CGG sequences the most. However, *in vivo* results show the opposite trend is true – the expansion frequency for the CAG triplet repeat sequences have been shown to be higher than any of the other triplet repeat sequences observed.[18] This is possibly due to the efficiency of the DNA repair system that more easily recognizes the CGG and CTG mismatches than the CAG mismatch.[168] It is possible that either the DNA repair system is very efficient at detecting certain types of DNA lesions, or that there is a different correlation between the thermodynamics and the expansion mechanism. Such a correlation may not be directly linked to the stability of the hairpin structure, but rather the probability that the end-frayed 3' end is in the correct orientation for stable hydrogen bonding and nucleation to occur between the frayed 3' end and the parent DA strand. Such probabilities could be dependent upon the underlying thermodynamics of the hairpin-parent DNA interaction.

Comparison of the extreme ends of the stability spectrum reveals that the stability of the CGG oligomers is highest and is driven by the favorable enthalpy. The entropy term for this oligomer is unfavorable relative to the CCG entropy and indicates that the CGG structure is more ordered than the CCG structure. There is more conformational freedom associated with the CCG oligomer than the CGG oligomer. While this conformational freedom may increase the probability of a CCG finding a matching CGG on the parent strand, it is improbable that the orientation may be correct, or that the base-pair once formed will be stable enough to continue propagating down the DNA chain due to the conformational freedom of the CCG hairpin. By similar analogy, ordered structures such as the CGG hairpin may make it relatively easier for the correct hydrogen bonding orientation to be achieved due to the limited freedom about the 3'-end, and once formed may allow the newly formed base pair to propagate along the DNA chain.

As shown within this work there is also a length component associated with the triplet repeat sequence thermodynamics (4.3.2.1). With all strands studied, as the longer the length of the oligomer, the more stable the oligomer structure became. This suggests a length dependent relationship between the probability for further expansion and the stability of the hairpin structure that is formed during replication. The longer the hairpin becomes the more stable becomes, and the more likely that the hairpin will remain as a stable structure prior to the reassociation of the DNA_{pol} complex. However, there appeared to be a practical limit to the stability gains achievable by the "simple" addition of additional triplet repeats units. Within this work (4.3.2.1) it has been shown that the stability, enthalpy, and entropy of the triplet repeat oligomers tended to reach a plateau at

approximately 45 nucleotides (or approximately 21 base pairs in the stem). The *in vitro* thermodynamic behavior of the triplet repeat oligomers predicts that *in vivo* the hairpins will reach a point beyond which only slight gains in stability can be achieved. This suggests that either the hairpins assume multiple but smaller structures to maintain the stability or that the cooperative unit of the hairpin is somewhere around 45 nucleotides. In either case, the "stability plateau" as it has been referred to in this dissertation allows the formation of a quasi-stable structure that is neither so unstable that it can readily decay back to a single-strand species and reanneal to the parent strand but not so stable that further structure formation is impossible. If the stability of the hairpin were too high, such as that observed for duplex DNA, then there would be no localized denaturation/reannealing (a.k.a. "breathing") within the molecule, and no opportunity for the formation of slip-structures. In such a scenario, denaturation of the hairpin could not occur without the aid of a DNA processing enzymes. In such a quasi-stable state, the hairpin is free to slip or slide to through localized denaturation events and allow for the creation of alternative loop size, stem length or even structures depending upon the ambient solution conditions.

In addition to the thermodynamic component this work also shows that there is also a kinetic component that factors into the prediction of whether or not a triplet repeat hairpin will lead to expansion *in vivo*. In a simplistic view, it has been proposed that reassociation of the DNA_{pol} complex and sequence expansion is a question of timing.[29] If the hairpin reanneals to the parent duplex prior to the association of the DNA_{pol} complex, then no expansion in triplet repeat number occurs because DNA synthesis has resumed

where the DNA_{pol} complex left off. If however, the hairpin does not reanneal, then association of the DNA_{pol} complex occurs downstream of the dissociation location, and reiterative DNA synthesis ensues. It has been shown both in the work of Gacy and in this dissertation that there is a length dependence to the reassociation kinetics.[29] Additionally, it has been shown, within this dissertation that there is a sequence context dependent to the reassociation kinetics. This context-effect is most likely related to the stabilities of the N.N mismatches.

The length dependence described within this thesis (4.4.1.1) suggests that the observed rate constants for the reannealing of long hairpin oligonucleotides to the complementary oligomer occurred were lower than the rate constants observed for short oligomer pairs. This observation, while not new (see Gacy, McMurray [29]) was interesting because it is inconsistent with what is known for the reannealing rates of genomic duplex DNA by C_0t analysis. In C_0t analysis a segment of DNA is first denatured to single strands and then allowed to reanneal to genomic DNA that contains the complementary DNA segment. The progress of this reaction is monitored by measuring the absorbance as a function of time and a plot of the initial concentration (C_0) versus time is constructed. The time it takes to reach one-half of the starting concentration gives an indication of the complexity of the DNA segment under consideration. The C_0t rates will increase the lower the frequency of the DNA complement within a genome and with increasing sequence complexity or uniqueness. The harder it is for a segment of DNA to find a complementary sequence out a few hundred thousand possibilities, due to low copy number or high complexity, the longer it takes for annealing to occur. Triplet Repeat

DNA is redundant and in the kinetics experiments described above did not have a large subpopulation of noncomplementary sequences. Therefore, the reassociation rates should have been independent of the length. This thesis demonstrates that the observed rate constants for short oligomers are higher than the observed rate constants for long oligomers. This relationship remains the same between both sets of oligomers either as (CAG)n/(CTG)n or (CCG)n/(CGG)n. In a cellular sense, this means that it is probable, that once a hairpin forms on the daughter DNA strand and finds a nucleation site, it is likely that longer the hairpin is, the slower it will reanneal back to the parental duplex. In terms of disease pathology this makes it more likely that as the length of the reannealing hairpin increased, that the DNA_{pol} complex would reassociate at a position downstream of the dissociation point.

It was shown from the Eyring plots within this dissertation that short oligomers reanneal with less disorder and less bond breakage in the transition state than long oligomers and from the Arrhenius plots with less activation energy than long oligomers. Disorder in the transition state is the result of conformational freedom of the hairpin, that includes the *anti*/*syn*/wobble gyrations of the mismatch basepairs, bending about the main axis of the DNA hairpin, and the motion of the free 3'end. With increased entropy in the transition state comes decreased ability for the free 3'end to obtain a nucleation partner sequence. This could be due to a problem with finding the correct sequences, or with maintaining the correct orientation for hydrogen bonding, or perhaps even to the self-association kinetics. The daughter hairpin to parent DNA process is one that is in competition with the self-association of the free 3'-end with the semi-complementary sequence of hairpin

stem. Increased conformational freedom, such as that observed in the transition state of long oligomers may translate into increased self-association kinetics, due to the fact that the complementary basepairs within the hairpin fork are closer than the complementary basepairs of the parent DNA. Should nucleation occur with the fork, then the reverse process- self association will occur and slow down the competing process where the free 3'end is attempting to anneal to the parent strand. The *in vivo* consequence of these observations is that short oligomers reanneal faster because the free 3' ends of the hairpin is can find a nucleation site faster, and there is less probability, due to the short length, for competing self-assembly to occur once the hairpin opens.

It has been shown within this work (4.4.1.1) that, like the thermodynamic component, sequence dependence also factors into the reassociation kinetics. On the basis of the Na^+ effect studies, it was shown that the oligomers undergo transition states that were dependent upon the sequence under consideration. The (CCG)n/(CGG)n oligomers reassociated with a more cruciform like intermediate, where as the (CAG)n/(CTG)n oligomers reassociated with a more single-strand like intermediate. Based upon Arrhenius plot data the (CCG)n/(CGG)n oligomers reassociated with faster kinetics than the (CAG)n/(CTG)n oligomers due to more bond breakage and more entropy in the transition state (Eyring plot). At the very least, the *in vitro* data suggests that *in vivo*, the (CCG)n/(CGG)n hairpins with the parent DNA strand through a more ordered intermediate than do the (CAG)n/(CTG)n oligomer pairs. It is believed on the basis of what has been presented here, that a cruciform intermediate makes it easier for the correct

alignment of the initial nucleation site to be achieved, and, in addition makes it easier for all downstream nucleotide dot be in the correct alignment of for annealing.

In summation, the probability for expansion of triplet repeat DNA sequences is dependent on whether or not the hairpin structure formed during replication remains prior to the reassociation of the DNA_{pol} complex. This dissertation suggests that the probability for the stable hairpin formation is increased with increasing length of the hairpin sequence and sequence context (G>T>A>C). The stability of the hairpin appears to be driven by the stability and thermodynamics of the mismatch and reaches a plateau at approximately 45 nucleotides. This work also suggests that the probability for expansion is also related to the reassociation kinetics. End-fraying at the 3' end of the hairpin introduces the possibility that the 3'end will nucleate with the parent DNA strand and then propagate a realignment of the nucleotides downstream of the nucleation site. At the same time there is a competing self-association process occurring with the hairpin nucleotides located closest to the hairpin fork. The rate of the overall "reaction" is dependent on the rate of parent DNA versus self-association. These rates were observed to be affected by the sequence context, and the sequence length. Long oligomers associated with slower kinetics than did short oligonucleotide pairs. Likewise, the (CAG)n/(CTG)n oligomer pairs associated with slower kinetics than did the (CCG)n/(CGG)n oligomer pairs due to differences in the intermediates or transition states for the reassociation process.

6 REFERENCES AND EXPLANATORY NOTES

[1] Mathews, C. K.; van Holde K. E.; **(1990)** *BIOCHEMISTRY*, Benjamin Cummings Publishers, ISBN 0-8053-5015-2; 28:997-998.
[2] Jeffreys, A.J.; Wilson, V.; Thien, S.L.; **(1985)** *Science* 314:67-73.
[3] Blackburn, G.M.; Gait, M.J.; **(1996)** *Nucleic Acids in Chemistry and Biology*, Oxford University Press, ISBN 0-19-963533-1; 5:214-217
[4] Freifelder, D.; **(1987)** *Molecular Biology 2nd Ed.*, Jones and Bartlett Publishers, ISBN 0-86720-069-3; 4:103-105.
[5] Paulson, H.L.; Fischbeck, K.H.; **(1996)** *Annu. Rev. Neurosci.*, 19:89, figure 1.
[6] Chen, F-M.; **(1992)** *Biochem.*, 31:3769-3776.
[7] Mathews, C. K.; van Holde, K. E.; **(1990)** *BIOCHEMISTRY*, Benjamin Cummings Publishers, ISBN 0-8053-5015-2; 28:997-998.
[8] Because there are three nucleotides that specify the repeat, the total possible combinations, given 4 different nucleotide bases is 4^3 or 64 different combinations. This is, incidentally, the same number of codons in the genetic code.
[9] Mandel, J-L.; **(1989)** *Nature Genetics*, 4:8-16.
[10] Pearson, C. E.; Sinden, R. R.; **(1996)** *Biochem.*, 35:5041-5053.
[11] Reddy, P.S.; Housman, D.E.; **(1997)** *Curr. Opin. Cell Biol.*, 9:364-372.
[12] Gacy, A.M.; Geollner, G.; Juranic, N.; Macura, S.; McMurray, C.T.; **(1995)** *Cell*, 81:553-540.
[13] This gene has been named the "Huntingtin gene" because the true cellular function of the protein product is unknown to date.
[14] Paulson, H.L.; Fischbeck, K.H.; **(1996)** *Annu. Rev. Neurosci.*, 19:70-107.
[15] Zheng, M.; Huang, X.; Smith, K.; Yang, X.; Gao, X.; **(1996)** *J. Mol Biol.*, 264:323-326.
[16] Sutherland, G.R.; Richards, R.I.; **(1995)** *Proc. Nat. Acad. Sci. USA*, 92:3636-3641.
[17] Hannan, A.; **(1989)** *Exp. Pharmacol. Phys.*, 23:1015-1020.
[18] Peyret. N. P.; Seneviratne, A.; Allawi, H.T.; SantaLucia, J. Jr.; **(1999)** *Biochem.*, 38:3468-3477.
[19] LaSpada, A.R.; Paulson, H.L.; Fischbeck, K.H.; **(1994)** *Ann. Neurol.*, 36:814-822.
[20] Verkek, A.J.; Pieretti, M.; Sutcliffe, J.S.; Fu, Y.H.; Kuhl, D.P.; **(1991)** *Cell*, 65:905-914.
[21] Kremer. B.; Goldberg, P.; Andrew, S.E.; Tehilman, J.; et.al.; **(1991)** *Science*, 252:1711-1714.
[22] Siomi, H.; Siomi, M.C.; Nussbaum, R.L.; Dreyfuss, G.; **(1993)** *Cell*, 74:291-298.
[23] Mitas, M.; **(1997)** *Nuc. Acids Res.*, 25:2245-2253.
[24] Fry, M.; Loeb, A.L.; **(1994)** *Biochem.*, 91:4950-4954.
[25] Fry, M.; Nadel, Y.; Weisman, P.; **(1995)** *J. Biol. Chem.*, 270:28970-28977.
[26] Chen, X.; Maraippan, S.V.; Catasti, P.; Ratcliff, R.; Moyzis, R.K.; et al.; **(1995)** *Proc. Natl. Acad. Sci.*, 92:5199-5203
[27] Nadel, Y.; Weisman-Shomer, P.; Fry, M.; **(1995)** *J. Biol. Chem.*, 270:28970-28977.
[28] Petruska, J.; Arnheim, N.; Goodman, M.F.; **(1996)** *Nuc. Acids. Res.*, 24:1992-1998.
[29] Gacy, A.M.; McMurray, C.T.; **(1998)** *Biochem.*, 37:9426-9434.
[30] Gellibolian, R.; Bacolla, A.; Wells, R.D.; **(1997)** *J. Biol. Chem*, 272 :16793-16797.
[31] Bacolla, A.; Gellibollian, R.; Shimizu, M.; et al.; **(1997)** *J. Biol.Chem*, 272:16783-16792.
[32] The DNA_{pol} complex is composed of helicase, primase, and polymerase.
[33] This phenomenon is the DNA equivalent of twisting an elastic rubber band on one end while keeping the other end fixed. Eventually the rubber band becomes so twisted that no further twisting can be accomplished without first relieving the accumulated tension or mechanical failure of the rubber band
[34] This is called a solitary oligomer to denote the absence of a complementary DNA strand that can form Watson-Crick type of hydrogen bonding. Although these oligomers are literally single strands with some associated structure, it is not to be confused with "single-strand" DNA that commonly refers to denatured or random coil DNA. Single-strand DNA has no structure, solitary DNA (in the context used for this dissertation) does have structure.

[35] Johnson, W.C.; **(1994)** *"CD of Nucleic Acids" in CD of NUCLEIC ACIDS, PRICIPLES AND APPLICATIONS*, K. Nakanishi, N. Berova, R.W. Woody, eds., VCH Publishers, ISBN 1-56081-618;; 19:523-540.
[36] Marky, L.A.; Breslauer, K.J.; **(1987)** *Biopolymers*, 26:1601-1620.
[37] This has also been called the hypochromic effect in reference to the fact that the native DNA absorbance is lower than the absorbance obtained for random coil DNA. "Hypo" or "hyper" chromicity depends upon the frame of reference used. Hyperchromicity if the reference point is the native DNA, with a lower starting absorbance, Hypochromicity if the reference point is the random coil DNA with a higher starting absorbance.
[38] Avizonas, D.Z.; Kearns, D.R.; **(1994)** *Biopolymers*, 35:187-200.
[39] The wavelength of maximum absorbance, 260 nm, is used for all analytical separations to maximize sensitivity. The off-peak wavelength, 280 nm, is used for preparative separations due to the large sample sizes typically injected onto the HPLC column.
[40] Standard HPLC practice is to either degas the buffers and solvents just prior to use or better yet, to use a constant helium sparge. Solution off-gassing (release of dissolved gasses) is one of the enemies of good HPLC. Release of dissolved gases will usually occur at the detector, as the eluent goes from high pressure to low. This tends to introduce noise, or spikes that will decrease chromatographic sensitivity. In addition, off-gassing can also occur on the column with ruinous effects.
[41] The water bath for the thermoelectric heating/cooling unit, also called a Peltier after the inventor, must be set for 20 °C any time heating and cooling control is desired. The Peltier device uses the water bath for heat dissipation. Failure to turn on the water bath can damage the Peltier.
[42] Disk Operating System, or "DOS" is a command-line interface between the user and the computer that predates Windows. There are no user-friendly windows in DOS, and it is very easy to make mistakes with this interface. Below are some of the rudimentary commands you will need to know in order to manipulate your files. Remember that DOS filenames can be no more than 8 characters long, with a three character extension/suffix.

Action	Command Structure	Command Line Example
copy files	copy_ sourcedrive\subdir\file.ext_destdrive\subdir\file	copy C:\data\tony\SVD.smo a:\
copy file, with wildcard	copy_sourcedrive\subdir\file*.ext_destdrive\subdir\file	copy C:\data\tony\SVD*.* a:\
copy file, with wildcard	copy_sourcedrive\subdir\file?1.ext_destdrive\subdir\file	copy C:\data\tony\SVD09??.smo a:\
copy file, with wildcard	copy_sourcedrive\subdir\file?1.*_destdrive\subdir\file	copy C:\data\tony\SVD09??.* a:\
copy file, with wildcard	copy_sourcedrive\subdir\??file1.*_destdrive\subdir\file	copy C:\data\tony\???SVD.* a:\
copy file, with wildcard	copy_sourcedrive\subdir\??file1.*_destdrive\subdir\file	copy C:\data\tony\????SVD.smo a:\
change dir, downward	cd_drive\subdir	cd c:\tony
change dir, upward to root directory	cd \	cd \
change dir, upward	c:\	C:\
directory	dir	C:\data\tony\dir
directory,pause	dir/p	C:\data\tony\dir/p
make directory		mkdir C:\data\jaime
move files	move_sourcedrive\subdir\file.ext_destinationdrive\subdir\file	move C:\tony\SVD.smo a:\
delete files	del_sourcedrive\subdir\filename	del c:\tony\svd.smo

NOTES:

copy: In the example above, the copy command will copy the file called "SVD.smo" from the tony subdirectory of the data subdirectory of the C drive, onto the A drive. In this case the A drive is a 3.5" diskette. The asterisk (*) wildcard is used for any alpha-numeric value in the position the wild card occupies and any position thereafter! For example the command "copy c:\tony\data*.smo" will copy all files with the first four positions containing the word "data" and ANY character in the last 4 positions of the file name. The question mark (?) wildcard is position sensitive. It instructs the computer to copy only those files that meet the wildcard and name specifications you give the machine. For example, the command "copy c:\tony\data?.smo" will only copy those files with data in the first four positions of the filename, and any character in the fifth position of the filename.
move: The move command moves files from one location of the disk drive directory structure to another or from one disk to another. This is the DOS equivalent of cutting and deleting in Windows. Because of this, use this function with extreme care, as it is quite easy to move a file and lose track of where it was moved to. There is no "find" function in DOS. I recommend this command for experienced users only.
delete: This command permanently deletes files from the computer. Use this command with extreme caution. There is no easy way (if any?) to recover the files from the hard drive once they have been deleted. I recommend this only for experienced users only.
[43] The NanoII DSC is a power-compensation calorimeter.
[44] The dessicator is used without dessicant to reduce sample condensation. When under vacuum, water will have a greater tendency to evaporate from the solution being degassed, resulting in concentration of the sample solution. Do not accelerate this process by adding dessicant to the dessicator. Glass dessicators make good vacuum degassing devices – they are inert, can withstand high vacuum, can maintain a vacuum for days and require little maintenance.
[45] Initially CruChem synthesis reagents were used exclusively, until we began experiencing batch consistency problems. Since then (Sept. 2000) all syntheses have utilized reagents from Glen research, except as noted for the ammonia and acetonitrile.
[46] The moisture content should be at or below 0.001%. This is currently Aldrich catalog number 27,100-4.
[47] The bottles were dried at 100 °C in a drying oven for a minimum of 24 hours prior to use. Rubber septa were immediately inserted upon removal from the oven and the bottles were purged and simultaneously cooled under argon prior to acetonitrile transfer.
[48] The recommended setting for the 380B synthesizer is "60 psi unless otherwise noted by the service/installation representative." I found that optimal flow rates were obtained with the pressure regulator set for 70 psi and not the recommended 60 psi as per ABI spec. It should go without saying but...the flow rates for all reagent lines should be checked if for some reason the pressure setting is changed.
[49] The thermal mass the electrophoreses tank is low and modification of the electrophoresis tank was necessary to limit the temperature drift during a separation. The temperature of an unheated tank can decrease by as much as 20 – 30 degrees during a separation and can lower the extent of denaturation. The modification of the tank allows external control of temperature, and limits the loss of heat during the course of a separation.
[50] Tygon tubing is an inefficient heat-transfer medium. To maintain a constant temperature of 60 °C inside the tank, it was necessary to set the water bath to approximately 80 °C. A better design would have utilized glass tubing inside the electrophoresis tank as the heat-exchanger, with Tygon tubing serving as conduit in between the water bath and glass tubing. Unfortunately, my glass-working skills were not up to this task.
[51] Safety Note: The specified maximum temperature for the Hoeffer model SE600 heating core is 40 °C. However, many analytical UREA-PAGE gels were routinely performed with this electrophoresis unit at 60 °C without any visible damage to the seals or the core itself. It is possible however, that eventually the seals may fail at the elevated temperatures used for Urea-PAGE. As a safety precaution, the lack of conductivity between the buffer reservoir and the water bath should be confirmed with a volt-meter prior to each gel performed.
[52] To ensure accuracy, calibrate the spectrophotometer using a NIST calibrated neutral density filter(for detector response) and a NIST calibrated holmium oxide filter (for monochromator accuracy). When using different spectrophotometers, it is good practice to calibrate the new instrument with at least the neutral density filters so that differences in instrument response can be eliminated.
[53] Water should not be used to fill the bath because it will completely evaporate over the required 18 hour reaction time.

time. To be consistent with what has been proven to work, the deprotection time should be kept as close as possible to 18 hours.

[66] The oil used for this purpose should come from the surrounding oil bath. Do not use room temperature oil.

[67] Great care was taken to ensure that the solution levels, including the crude oligomer, the beaker oil level, and the bath oil level, were equivalent. If the oil level of the surrounding bath was too low relative to the oligomer solution level, incomplete thermal exchange will occur. If the oil level of the surrounding bath was too high with respect to the solution level of the crude oligomer, the tubes may tip-over and possibly introduce oil contamination. While it should go without saying, gloves should be worn when adjusting the level of the lab-jack. Because similar volumes (4-6 ml of DNA solution in ammonia) were used from "run-to-run", the lab-jack only had to be adjusted once to be within 0.5 cm of the correct level.

[68] The synthesis reaction proceeds from the 3' direction to the 5' direction, with the last DMT group on the 5' end remaining on the synthesized oligomer. Thus a 30-mer should have (30-1) or 29 DMT peaks. The general formula for a DMT-on synthesis is tirtyls = (n-1) where n = oligomer length.

[69] Occasionally, due to idiosyncrasies of the 380B synthesizer, the observed step yields would momentarily decline over 1 or 2 synthesis cycles but then recover. Some of these declines were due to the failure of the synthesizer to fully push all the trityl solution through the waste line, and not due to a true decrease in step yield. These oligomers were considered as normal, provided the trityl assay recovered within one or at most two cycles downstream.

[70] A flow rate of 1 drop/sec roughly corresponds to a flow rate of 1 –2 ml/min. For this and all subsequent steps, the flow rates should be between 1 –2 ml/min. At no time during this procedure should the column be allowed to run dry because this will decrease yields.

[71] Typically, in RPHPLC the failure sequences are removed by elution with 5-10% v/v acetonitrile in water. The failure sequences are less hydrophobic than the DMT-on oligomer, making these sequences more likely to elute from the column in low levels of acetonitrile. With PPII columns however 10% ammonia is used to accomplish the same separation. Ammonia acts as an ion pairing reagent to the negatively charged phosphate backbone of the DNA strand and helps to stabilize secondary structure formation. This in turn decreases the hydrophobic interaction between the oligomer and solid support, and makes the failure sequences less likely to adhere to the column.

[72] The Glen Research specification for this and all other wash steps with a PPII column is 4 ml. The wash volumes used throughout this procedure are considerably larger and the final wash volume is larger by almost a factor of 10. The dry-bed volume, measured directly through cartridge-dissection, is approximately 500 µl. Therefore a 4 ml wash volume is, assuming a complete change of solution (1:2 dilution) every 500 µl, an approximate change of 2^8 or 1:256 dilution factor. A 10 ml wash volume produces a 2^{20} or $1:10^6$ dilution factor. A 30 ml wash volume produces a 2^{60} or $1:10^{18}$ dilution factor. Large wash volumes are used is to ensure complete and exhaustive removal of prior salt ions. The use of exhaustive washing at the end of the PPII step eliminates the need for downstream dialysis, and avoids the loss of oligomer occasionally observed with dialysis procedures.

[73] The Econo-Pump was not used for the column preparation due to incompatibility of the pump tubing to acetonitrile.

[74] While it is possible to load the crude oligomer solution onto the PPII column with the cartridge connected to the Econo-Pump system it was not attempted for several reasons. First, the flow rate, must be slower than the 1- 2 ml/min flow rate possible with the Econo-Pump. While the solution could be pushed through the column manually with the absorbance detector down-line, it is not wise to do so because of chemical incompatibility between the silicon rubber tubing and the concentrated ammonia solution. Second, DNA cross-contamination resulting from user error, small droplets within the tubing lines, or DNA absorption to the tubing is high when loading the crude DNA solution in such a manner. For these reasons, the Econo-Pump system is used only to monitor the washing steps of the purification process so as to avoid cross-contamination of the final product. The DNA being purified is never allowed to come in contact with the downstream end of the purification system. Low purification yields can usually be traced back to two sources of error – the DNA did not absorb to the PPII column, or there was a catastrophic loss of DNA during one of the washing/processing steps. The Econo-Pump allows the user to determine if the latter situation occurred.

[75] It is necessary to remove the Econo-Pump from the effluent stream so that a controlled flow rate of 5 ml/min can be achieved. The detector however is still on-line downstream of the PPII column to allow monitoring of the detritylation reaction.
[76] The μ-Bondapak RPHPLC column contains a silica solid support. Silica solid supports are susceptible to base-hydrolysis and will eventually lead to poor chromatographic separations.
[77] DNA is acid-labile, so precautions must be taken during the lyophilization process to ensure that the pH remains basic after the ammonia is removed.
[78] For most of my oligomers I used the Savant Speed Vac because multiple samples could be processed at one time. When using the Speed-Vac for this procedure use the caps-on approach (bottle-caps on with hole in the center) due to the high possibility of bumping. Alternatively, the oligomer solution could be lyophilized with a RotaVap concentrator. Prior RotaVap procedures describe the use of triethylamine and ethanol. Ethanol is used as an azeotrope during lyophilization. An azeotrope will hydrogen-bond to the water molecules, and, because of the higher vapor-pressure of the ethanol, make it easier to remove water from the solution. Please consult the S. Marotta, or A. Anathayan dissertations for the details of this procedure. The ethanol/triethylamine protocol was not used for the Savant Speed Vac procedures because of problems observed in the dissolution of highly hydrophobic G/C rich sequences.
[79] With experience, it was found that the analytical injection of oligomer solution prior to the preparatory separation was helpful, but not necessary. The analytical run saves time because DNA samples with high levels of failure sequences can be eliminated. However the analytical run, while bearing a resemblance to the prep run, is a poor predictor of the start and stop points for fraction collection due to differences in injection scale and the gradient slope. Minor or barely detectable peaks on the analytical run became "mountainous peaks" during the preparatory separation. This reduced the procedure of fraction collection to an "on-the-fly" process based upon experience. Rule of thumb – if in doubt collect the peak. One can always discard the collected fraction later when the full chromatogram is in view.
[80] Marotta, S. ; Dissertation Seton Hall University.
[81] This is a new gradient, based upon the gradients developed by S. Marotta, to separate the CGG sequences. The original naming convention was retained.
[82] A microcentrifuge was used to hasten the separation process if the solution volume during the ether extraction was less than 2 ml. The samples were spun at 14000 rpm for 2 minutes at room temperature.
[83] Occasionally, oligomers were purified a second time using this gradient even though the prior RPHPLC, AXHPLC or gel analysis indicated that the oligomers were pure enough – this judgment call is solely up to the practitioner.
[84] Remember that the specified pore size is a range of pore sizes and not an exact number.
[85] Dialysis occurs for approximately 6 hours before the dialysis solution is refreshed with the next buffer change. Most processes will reach equilibrium by this time.
[86] A sample calculation: A 1 ml DNA sample in 1 molar sodium chloride was dialyzed against two 1 liter volumes of buffer. What is the salt exchange factor?

$$SEF = (1000/1)(1000/1) = 10^6$$

The original salt concentration was diluted 1 million-fold. If the original salt concentration was 1 molar, then the final concentration is 1molar/1,000,000 = 1 μM salt. This salt concentration is too low to be of any significance and can be considered as effectively removed.
[87] The molecular mass of monobasic sodium phosphate is 120, the mass of CAG12 is 3400, the pore size selected was 1000 MWCO.
[88] Gloves should be worn when handling the dialysis tubing for two reasons. First bare skin contains numerous bacteria, proteases, DNAses, and RNAses – the very things being removed by the boiling and buffer exchange procedures. Second, contaminating proteins and other cellular debri that can rub off of bare skin may clog the pores of the membrane.
[89] Gentle stirring! The rotation of the stir bar should be fast enough to keep the dialysis bag in constant motion (to ensure rapid removal and dispersion of any ions that cross the membrane) yet not so fast that the dialysis bag is drawn down, by vortex action, to the stir bar. Contact with the stir bar can damage the tubing and may result in the catastrophic loss of contents. This is the equivalent of a piñata being hit by a baseball bat – sooner or later the piñata will break.
[90] Arghavani, M.B.; Romano, L.J.; **(1995)** *Anal. Biochem.*, 231:201-209.

[75] It is necessary to remove the Econo-Pump from the effluent stream so that a controlled flow rate of 5 ml/min can be achieved. The detector however is still on-line downstream of the PPII column to allow monitoring of the detritylation reaction.

[76] The μ-Bondapak RPHPLC column contains a silica solid support. Silica solid supports are susceptible to base-hydrolysis and will eventually lead to poor chromatographic separations.

[77] DNA is acid-labile, so precautions must be taken during the lyophilization process to ensure that the pH remains basic after the ammonia is removed.

[78] For most of my oligomers I used the Savant Speed Vac because multiple samples could be processed at one time. When using the Speed-Vac for this procedure use the caps-on approach (bottle-caps on with hole in the center) due to the high possibility of bumping. Alternatively, the oligomer solution could be lyophilized with a RotaVap concentrator. Prior RotaVap procedures describe the use of triethylamine and ethanol. Ethanol is used as an azeotrope during lyophilization. An azeotrope will hydrogen-bond to the water molecules, and, because of the higher vapor-pressure of the ethanol, make it easier to remove water from the solution. Please consult the S. Marotta, or A. Anathayan dissertations for the details of this procedure. The ethanol/triethylamine protocol was not used for the Savant Speed Vac procedures because of problems observed in the dissolution of highly hydrophobic G/C rich sequences.

[79] With experience, it was found that the analytical injection of oligomer solution prior to the preparatory separation was helpful, but not necessary. The analytical run saves time because DNA samples with high levels of failure sequences can be eliminated. However the analytical run, while bearing a resemblance to the prep run, is a poor predictor of the start and stop points for fraction collection due to differences in injection scale and the gradient slope. Minor or barely detectable peaks on the analytical run became "mountainous peaks" during the preparatory separation. This reduced the procedure of fraction collection to an "on-the-fly" process based upon experience. Rule of thumb – if in doubt collect the peak. One can always discard the collected fraction later when the full chromatogram is in view.

[80] Marotta, S. ; Dissertation Seton Hall University.

[81] This is a new gradient, based upon the gradients developed by S. Marotta, to separate the CGG sequences. The original naming convention was retained.

[82] A microcentrifuge was used to hasten the separation process if the solution volume during the ether extraction was less than 2 ml. The samples were spun at 14000 rpm for 2 minutes at room temperature.

[83] Occasionally, oligomers were purified a second time using this gradient even though the prior RPHPLC, AXHPLC or gel analysis indicated that the oligomers were pure enough – this judgment call is solely up to the practitioner.

[84] Remember that the specified pore size is a range of pore sizes and not an exact number.

[85] Dialysis occurs for approximately 6 hours before the dialysis solution is refreshed with the next buffer change. Most processes will reach equilibrium by this time.

[86] A sample calculation. A 1 ml DNA sample in 1 molar sodium chloride was dialyzed against two 1 liter volumes of buffer. What is the salt exchange factor?

$$SEF = (1000/1)(1000/1) = 10^6$$

The original salt concentration was diluted 1 million-fold. If the original salt concentration was 1 molar, then the final concentration is 1molar/1,000,000 = 1 μM salt. This salt concentration is too low to be of any significance and can be considered as effectively removed.

[87] The molecular mass of monobasic sodium phosphate is 120, the mass of CAG12 is 3400, the pore size selected was 1000 MWCO.

[88] Gloves should be worn when handling the dialysis tubing for two reasons. First bare skin contains numerous bacteria, proteases, DNAses, and RNAses – the very things being removed by the boiling and buffer exchange procedures. Second, contaminating proteins and other cellular debri that can rub off of bare skin may clog the pores of the membrane.

[89] Gentle stirring! The rotation of the stir bar should be fast enough to keep the dialysis bag in constant motion (to ensure rapid removal and dispersion of any ions that cross the membrane) yet not so fast that the dialysis bag is drawn down, by vortex action, to the stir bar. Contact with the stir bar can damage the tubing and may result in the catastrophic loss of contents. This is the equivalent of a piñata being hit by a baseball bat – sooner or later the piñata will break.

[90] Arghavani, M.B ; Romano, L.J.; **(1995)** *Anal. Biochem* , 231:201-209.

[91] Preparative-scale anion exchange columns for HPLC do exist, but the associated price for such columns is beyond the budget of Seton Hall University. The capacity of most analytical columns is designed for analysis and not preparative work.

[92] This technique can easily be abused due to the ease with which some samples can be prepared and a spectrum recorded. This ease can lull novice users into a false sense of over-confidence that can lead to carelessness when using the spectrophotometer. When using the spectrophotometer one must always be aware of what is being measured, and how the sample was prepared. Samples with suspended precipitates will give erroneously high "absorbance" readings due not to a true absorption phenomena, but to light scattering. Similarly, preparation of samples with an out-of-specification pipettor or spectrophotometer, will lead to erroneous results.

[93] J. Waser; **(1964)** *Quantitative Chemistry,* Benjamin Cummings Inc. ISBN 0-8053-9503-2; 225-242.

[94] Unknown, I.M.; **(1995)** *Current Protocols in Molecular Biology.*

[95] Concentration critical experiments include: DSC, Kinetics, CD, UV-melts, and some gel electrophoresis experiments where comparison across samples was necessary.

[96] Non-critical concentration experiments include: Gel electrophoresis used for purity analysis, dialysis yield measurements, purification yield measurements, and routine concentration measurements. It can be argued that the measurements listed above certainly qualify for more accuracy than what was assigned. These measurements were used to derive approximate or "working" DNA concentrations with the assumption that if the DNA concentration needed to be more accurate the summed molar coefficient could be used. In most cases, the type of information sought in the assays listed was rough yield information within 10% RSD. Gel electrophoresis experiments for purity analysis were treated in the same manner. The concentration, be it 1 μg or 1.1 μg was not critical, as long as the DNA concentration was within some specified range. This specified range could easily be achieved with the averaged extinction coefficient. The averaged extinction coefficient offers one significant advantage over the other methods – convenience. With the average coefficient only two different numbers need to be memorized if the calculations are to be done on an on-site basis. Concentration and dilution calculations can be performed at the spectrophotometer without the need to refer to a table or spreadsheet. With the summed coefficient, many different numbers need to be memorized (one for each sequence under consideration) and are best left to spreadsheets.

[97] It may come as a surprise to some that the first step of sample preparation has nothing to do with the sample directly. Pipettor accuracy will vary. The pipette is the most critical component and also the weakest link in this measurement process. Calibrate the pipettor prior to use in this assay. To ensure greater accuracy, calibrate the spectrophotometer using a NIST calibrated neutral density filter(for detector response) and a NIST calibrated holmium oxide filter (for monochromator accuracy).

[98] Originally, the standard practice was to scan samples against a reference cuvette containing the same buffer solution as the sample solutions. This method however, only works when matched cuvette cells are used in both the sample and reference positions. Maintaining matched cells in a laboratory with multiple users can be difficult - invariably, the matched cells get misplaced, lost, dirty or broken. The problem became quite obvious one day when, attempting to measure concentration of a DNA solution in duplicate, the difference in absorbance between two samples was 0.1 A260 with reference cuvettes, and 0.01 without the reference cuvettes. A solution to this problem, albeit one that requires an extra measurement step, is to scan the sample cell containing buffer without a cuvette on the reference side (sample scan versus air), and then manually subtract the background absorbance for that cuvette from the spectrophotometric scan of the sample being measured.

[99] It was observed that some sequences of DNA, particularly those with high G content, were less soluble in PBS than other sequences. Sequences such as (CGG)n were more likely to form precipitates than sequences such as (CAG)n. This problem becomes particularly acute when, after storage at –20 °C, the sample is thawed for use. It was not uncommon to observe a precipitate at the bottom of the microcentrifuge tube. Eventually, from several hours to days, this precipitate would redissolve. Heating the samples to 90 –100 °C hastens the dissolution process.

[100] The sample must be thoroughly mixed prior to measurement. Simply leaving the sample on the lab-bench to thaw undisturbed will not evenly distribute the sample contents throughout the vial.

[101] Particulates must be removed for the purposes of spectral measurement because they will increase the noise of the spectrum, and may increase the apparent absorbance measurement. Centrifugation works reasonably well, but filtration through a 0.45 μm syringe filter is better. Filtration can lead to loss of

sample volume however, due to the void volume of the syringe filter assembly. For spectral scans, the effects of particulates are easier to identify – increase in noise, and an elevation in the absorbance near 400 nm. For single-wavelength absorbance measurements however, the effects of particulates are less obvious and can easily mislead the user with an apparently higher absorbance value.

[102] The sample is taken near the top of the solution to avoid drawing up the particulates that may have been pelleted in the previous step.

[103] Mixing of the sample is achieved by "pumping" action with the pipettor. Draw up approximately 100 µl of solution with a 200 µl pipettor and then rapidly expel the aliquot back into the cuvette. This pumping action is performed 5-10 times to thoroughly mix the contents.

[104] Bandwidth matters! The bandwidth is defined as the segment of spectrum passing through the sample at any given moment during a scan. To avoid loss of spectral information, the bandwidth should be the same as or slightly larger than the step size. The bandwidth also determines the resolution of the spectral scan – sharper peaks result from narrow bandwidths, whereas broader peaks result from wider bandwidths. However, at the same time, the noise levels will increase with narrow bandwidths, but decrease with wider bandwidths. Finally, the choice of bandwidth is also dependent upon the type of cuvette cell used for the measurement due to the optical layout for this spectrophotometer. Narrow bandwidths should be used for cells with narrow windows to prevent partial occlusion of the light beam as it enters the cell. The maximum bandwidth for a 2 mm wide cell is 2.0 nm, and for a 4 mm wide cell is 4.0 nm (personal communication from the Varian Service Representative). However, in practice, it was found that smaller bandwidths than those recommended worked better – 1.0 nm bandwidth for cells with 2 mm windows, and 2.0 nm bandwidth for cells with 4.0 mm wide windows.

[105] The overall shape of the spectrum provides information about the quality of the DNA. Highly pure DNA samples and sequences with high G-content tend to give relatively sharp peaks, because the absorbance near 280 nm is minimized.

[106] The DNA spectral scan should produce a single peak centered near 260 nm. Multiple peaks in this region indicate impurities. Multiple peaks, with maxima located near 250 nm and 270 nm (but not at 260 nm) indicate that the DNA sample has undergone depurination.

[107] Significant spectral noise at 400 nm implies particulates in the sample. Typically, optical density measurements, such as those used for cell growth, are performed at 600 nm. In this protocol, beginning at 400 nm allows shorter measurement times. In addition, not many compounds common to the DNA laboratory will absorb at 400 nm and, if they do, will produce an absorbance (as opposed to noise) spectrum.

[108] The A260/A280 ratio was "borrowed" from a routine measurement used by molecular biologists to determine DNA purity. Isolation of DNA from cellular sources can introduce contaminating RNA or proteins from copurification and, as a result, alter the measured absorbance at 260 nm. Ribonucleic acids will produce an A260/A280 of 2.0, whereas dsDNA has an A260/A280 of 1.8. Proteins absorb well at 280 nm and will reduce the A260/A280 ratio. However, the DNA used for all studies in this project was synthetically derived, and should not contain contaminating RNA or protein. The A260/A280 ratio in this instance indicates the presence of other organic molecules, such as exocyclic protecting groups, or trityl groups, that may have copurified with the DNA. These groups, will also absorb in the ultraviolet region of the spectrum near 280 nm.

[109] Cooper, T.; **(1977)** *The Tools of Biochemistry*, John Wiley and Sons, ISBN 0-471-17116-6, 6:194-200.

[110] The reason behind fixing the acrylamide:bis-acrylamide ratio at 19:1 was more practical than theoretical. Keeping the ratio constant enables preparation of gels from a single stock solution and lessens the chance that the wrong stock acrylamide solution is used to prepare gels.

[111] The resolution of a 20% polyacrylamide gel was better suited for short-mers, such as the 12-mer to 45-mer than it was for the 75-mers. For better resolution of the 75-mers, a 15% polyacrylamide gel was utilized. Rarely however, was it necessary to use 15% gels because reliable information regarding purity and size could be obtained for the 75-mer on a 20% gel. For critical applications, where resolution was required for both long and short oligomers, polyacrylamide gradient gels containing a pore gradient were used.

[112] Invitrogen Catalog, **(2001)** Invitrogen Inc., 11:254-255

[113] The other staining reagent used in this laboratory was "StainsAll" from Sigma Chemical Company. This staining solution stains both proteins and DNA, requires hours of staining to achieve the same sensitivity and allows the DNA bands to partially diffuse causing loss of band resolution.

[114] For most DNA stock solutions, one cannot directly dispense 0.03 or 0.003 μg of DNA. For these samples, dilute an aliquot of the stock solution 10-fold or 100-fold, then dispense from the dilutions. For example, if 0.03 μg of stock solution corresponds to 0.1 μl of DNA stock solution (too small to dispense accurately) then use 1 μl of a 1:10 dilution, or 10 μl of a 1:100 dilution.

[115] The samples are transferred directly from the water bath at 90 °C to the gel. At no time are the samples allowed to cool, other than the minimal time required to open the microcentrifuge aspirate the sample and load onto the gel. The remaining samples should stay inside of the water bath.

[116] Electrophoretic bands tend to diffuse or "spread-out" as migration distance increases. Diffusion can result in a distortion of the electrophoretic pattern, rendering purity analysis difficult or impossible. The tendency for diffusion can be minimized if sample buffer is placed in all empty lanes not containing a DNA sample. The sample buffer "crowds" the available space in the gel, forcing the sample DNA bands to stay more or less within lane.

[117] The use of glass trays should be avoided because Syber-Green I dye adheres to glass. Molecular Probes, Syber-Green I product description/insert

[118] The DNA bands will diffuse out of and within the gel resulting in a loss of resolution and a decrease in staining signal. Use the minimum time necessary.

[119] For optimum data interpretation, all electrophoresis gels should have well-defined bands, migrating evenly across the gel, with bands at defined locations. Gels with slow edge-migration bands relative to those in the center, will have an overall "smiling" appearance and make assignment of the correct oligomer size difficult. Distortions in migration patterns can occur for one of many reasons – the ionic nature of the sample was significantly different than that of the gel, uneven heating/cooling occurred during the electrophoretic separation, or bubble formation at the bottom of the gel cassette.

[120] Johnson, W.C.; **(1992)** *Methods in Enzymology*, 210:426-447.

[121] In this case slower is better. The object of this procedure is to allow the DNA as much time as possible to achieve the equilibrium structure upon renaturation. The slow cooling rate is accomplished on the heat-block, by simply heating up the specified temperature then shutting off the power to the heat block. The block will slowly cool to room temperature over the next 1-2 hours due to the thermal inertia of the block.

[122] Circular dichroism spectrophotometry is dependent on DNA structure. The DNA structure is dependent both on the solution conditions and the manner in which the DNA samples were prepared. Failure to rigorously follow sample treatment protocols from sample to sample within an experimental data set can result in alternative DNA structures that do not give consistent spectra. This is particularly critical with quadruplex-forming DNA sequences. The high heat, slow cool and 48-hour equilibration was necessary to allow the sequences sufficient time to self-assemble.

[123] Recall that the sieving effect of gel electrophoresis is related to the pore size of the gel, and the pore size is related to the percent acrylamide concentration, %C, and percent crosslinking, %T. Because the percent cross-linking (%T) was held constant the pore size and therefore the sieving effect, was manipulated through the acrylamide concentration. The sieving effect becomes stronger as the DNA migrates from the top to the bottom of the gel.

[124] The samples are heated to 90 °C to first denature the samples and remove any "spurious" DNA secondary structures that are the result of kinetically-based phenomena or storage conditions. The samples are slowly equilibrated to room temperature to allow the DNA to assume the most stable, equilibrium structure.

[125] The addition of glycerol to the acrylamide solutions increases both viscosity and density of the solutions. The increased viscosity minimizes mixing of the gel solution by reducing the turbulence at the gel interface. The density increase also allows efficient layering of the of the two acrylamide solutions and limits diffusion of the acrylamide prior to polymerization.

[126] The specified concentrations of ammonium persulfate and TEMED allow about 30 minutes of working time prior to polymerization. Inexperienced users, who need more time to pour a gel, should lower these amounts.

[127] Throughout this procedure I have reiterated the concept of 37 °C pre-equilibration. All solutions or equipment that come in contact with the sample (exception pipette tips) must be equilibrated at 37 °C to avoid thermally-induced structural variations. Hence the sample buffer, the electrophoresis buffer, the electrophoresis tank, the water baths, etc, are all set for and equilibrated at 37 °C prior to use with sample. This rule can be ignored however, once the electrophoretic separation is complete (ie the staining procedures).

[128] The use of glass trays should be avoided because the Syber-Green I dye adheres to glass and will decrease sensitivity. Molecular Probes, Syber-Green I product description/insert.
[129] The product insert is somewhat vague about this topic – Like ethidium bromide, Syber-Green I dye may be somewhat photosensitive. It is therefore good practice to cover the staining tray during incubation.
[130] The DNA bands will diffuse out of and within the gel resulting in a loss of resolution and a decrease in staining signal. Use the minimum time necessary.
[131] For optimum data interpretation, all electrophoresis gels should have well-defined bands, migrating evenly across the gel, with bands at defined locations. Gels with slow edge-migration bands relative to those in the center, will have an overall "smiling" appearance and make assignment of the correct oligomer size difficult. Distortions in migration patterns can occur for one of many reasons – the ionic nature of the sample was significantly different than that of the gel, uneven heating/cooling occurred during the electrophoretic separation, or bubble formation at the bottom of the gel cassette.
[132] A common mistake in sample-handling is to transfer the DNA samples to the cuvettes without any regard to the number artificially assigned to the cuvette. Often times, during the course of sample-preparation distractions will interrupt your work flow. By placing sample 1 in cuvette 1, sample 2 in cuvette 2 etc. there will be no doubts about where the sample is located upon resumption of work. In addition the use of numbered cuvettes in order avoids positioning mistakes when samples are placed in the spectrophotometer. As a rule of thumb – Sample1 –> cuvette1 -> position1 of the spectrophotometer cell holder. Maintaining the same cuvette with respect to spectrophotometer position also has another advantage – the optical performance of the cuvette(s) can be quickly and routinely monitored with time through comparison of a spectrophotometric scan of the buffer blanks.
[133] Teflon caps do not provide a secure seal at elevated temperatures because considerable pressure can build-up within a cuvette. Many samples have been lost by teflon caps that "pop" off the cuvette at elevated temperatures. This can result in sample loss either from evaporation or boil-over of the sample solution. Boil-over is particularly hazardous and can damage the circuit board of the Peltier located directly below the cuvette cell holder. The use of slightly oversized silicone-rubber stoppers (13D silicone-rubber Fisher Scientific Catalog number 09-704-1C) avoids most, but not all, of these problems. Proper insertion is important – be sure the bottom of the stopper protrudes into the cuvette space (is visible in the upper window).
[134] During the winter-time the ambient laboratory air is dry and condensation on the cuvette surfaces is less probable. Nitrogen-purging of the sample chamber was rarely used during this time. During the summer, late spring and early fall months (when relative humidity was high) nitrogen purging was utilized for all spectroscopic work below the ambient air temperature.
[135] Manning, G.S.; **(1978)** *Quart. Rev. Biophys.*, 11:179-246.
[136] Record, M.T.; Anderson, C.F.; Lohman, T.M.; **(1978)**, *Quart. Rev. Biophys.*, 11: 102-178.
[137] This procedure predates the DSC, and the assembly of the vacuum dessicator system. The Speed-Vac condensation trap is off to minimize water removal. I have been told by a former Speed-Vac sales representative (caveat here – the information comes from an unreferenced source) that the rate of lyophilization is proportional to the temperature difference between the sample chamber and the condensation trap. Therefore, to slow down water-vapor transfer the trap is off. The object here is to degas, not condense the sample.
[138] There are a few examples in the DNA literature of DNA sequences that melt without any hyperchromicity. Some quadruplex sequences are "notorious" for this effect.
[139] This procedure predates the DSC, and the assembly of the vacuum dessicator system. The Speed-Vac condensation trap is off to minimize water removal. I have been told by a former Speed-Vac sales representative (caveat here – the information comes from an unreferenced source) that the rate of lyophilization is proportional to the temperature difference between the sample chamber and the condensation trap. Therefore, to slow down water-vapor transfer the trap is off. The object here is to degas, not condense the sample.
[140] In practice, only three to four measurements are required for a reliable melting temperature. However, the UV-melt technique is fraught with so many problems arising from the sample, computer, or spectrophotometer, that is prudent to collect more data than needed rather than the bare minimum. Spectrophotometer problems include computer "crashing", problems with reaching the temperature set-points, condensation on the cuvettes, and random noise. Sample problems include hysteresis due to sample

pretreatment, bubble formation, noisy transitions, and optical interfaces (Schleirens) due to internal concentration changes.

[141] Use of the Peltier block robe set assumes that the thermal equilibrium between the block and the cuvette is more rapid than the rate of temperature change. This assumption is valid as long as the ramp rate is 0.5 °C/min or less.

[142] The external probe is inserted directly inside of the measurement cuvette to enable direct monitoring of the solution temperature. The external probe needs to be calibrated on a routine basis and thoroughly cleaned between samples.

[143] Langerman, N.; Biltonen, R.L.; **(1979)** *Methods in Enzymolog*, 302:261-286.

[144] Klotz, I.M.; Rosenburg, R.M.; **(1986)** *Chemical Thermodynamics* Benjamin Cummings Publisher, ISBN 0-8053-5501-4; 4:52-56.

[145] Leharne, S.A.; Chowdry, B.Z.; **(1998)** *Biocalorimetry*, J. Ladbury; B.Z. Chowdhry; Eds.; ISBN 0-471-97781-0; John Wiley and Sons, Chapter 12.

[146] Weidner, ; Sells ; **(1975)** *Elementary Physics*, ISBN 0-205-04647-9; Allyn and Bacon Inc. p.382.

[147] Privalov, G.P.; Privalov, P.L.; **(2000)** *Methods in Ezymology*, 323:31-62.

[148] The volume must be carefully selected. The sample cell has a volume of 299 µl, but a larger volume is necessary for the purposes of sample loading. Those familiar with the loading technique should use smaller volumes, about 750 µl whereas inexperienced users should use about 1000 µl.

[149] The buffer solution used for the final rinse of the capillary cells should be the same buffer solution used for the buffer blank (reference side of the DSC) and sample. Use of any other solution, such as water, can introduce subtle changes to the buffer composition, leading to inaccurate DSC quantitation. Droplets of the final rinse solution occasionally remain within the capillary cell prior to sample loading. Any solution remaining within the capillary cells after rinsing can change the composition of the sample or reference buffer upon filling the cell. It is therefore good practice to use the same buffer throughout, so that any droplets remaining after the final rinse changes only the sample concentration and not the buffer composition. It is this reason (and many others) that the sample concentration is measured both before and after a DSC experiment.

[150] This sample pumping or cycling back and forth procedure is described in the CSC manual for theNanoII DSC. There are clear and concise photographs of the loading procedures.
CSC Model 6100 Nano II DSC Operator's Manual, **(1998)** Section.4:40-76

[151] The equilibration process can take 45 minutes or more.

[152] Breslauer, K.; **(1994)** *Methods in Microbiology*, 26:347-372. Pay particular attention to pages 362-263 that describe the information that can be derived from comparison of van'T Hoff to calorimetric data.

[153] Similarly, a 12-bp duplex with a ΔHvh/ΔHcal ratio of 2 suggests that the cooperative unit is larger, perhaps as long as 24 bp long. I would however hesitate to make such claims on extrapolated data, unless there was a 24-bp duplex with a ΔHvh/ΔHcal ratio of 1 to substantiate this conclusion.

[154] Marky, L.; Breslauer, K.; **(1987)** *Biopolymers*, 26:1601-1620.

[155] CSC 6100 Nano II DSC Calorimeter Users Manual, 5:152-154.

[156] Noggle, J. H.; **(1989)** *Physical Chemistry 2nd Ed.* Harper Collins; ISBN: 0-673-39817; 10:540-546.

[157] Tinoco, I.; Sauer, K.; Wang, J.C.; **(1978)** *Physical Chemistry Principles and Applications in Biological Sciences*, Prentice-Hall Inc., ISBN 0-13-186545-5, 7:369-372.

[158] Here is a good example why the use of the A_{260} for expressing DNA concentration or amount can lead to erroneous results. The A_{260} for equimolar concentrations of CAG10 and CTG10 were 0.4 and 0.32 respectively. Use of equivalent A260 values (example A_{260} for CAG10 = 0.4 and CTG10 = 0.4) as the sole determinant for "DNA concentration" would have led to unequal molar DNA concentrations.

[159] This entails, a full Arrhenius-type activation energy analysis of 6 data points per temperature, over a range of 6 to 8 temperatures, at 6 different salt concentrations, for all 8 possible duplexes of the triplet repeat sequences under study – a total of 2304 kinetics traces that have to be manually processed for complete analysis. Assuming a maximum through put of 22 measurements/day over 10-12 hours translates into 104 days of kinetics measurements. At approximately one day to process 22 data files, means 420 hours or 52 days of processing time. This brings the grand total to 152 days to perform the comprehensive experiment under ideal conditions.

[160] Johnson, W.C.; **(1994)** in *Circular Dichroism, Principles and Applications*, Naknishi, K.; Berova, N.; Woody, R.W.; Eds; VCH Publishing, "CD of Nucleic Acids" 19:523-541.; specifically look at pages 534-535.

[161] Sheardy, R. D.; **(1991)** *Spectroscopy,*
[162] Marotta, S.P.; Tamburri, P.A.; Sheardy, R.D.; **(1996)** *Biochem.*, 35: 10484-10493; fig. 4.
[163] Manning, G.; **(1978)** *Accounts of Chem. Res.*, 12:443-449.
[164] Johnson, W.C.; **(1994)**, *CD of NUCLEIC ACIDS, PRICIPLES AND APPLICATIONS*, K. Nakanishi, N. Berova, R.W. Woody, eds., VCH Publishers, ISBN 1-56081-618, "CD of Nucleic Acids" 19:523-540. Pay particular attention t the point made on page 532, with figures 19.9, 19.5 and 19.10.
[165] Peyret, N.; Seneviratne, A.; Allawi, H.; SantaLucia Jr. J.; **(1999)** *Biochem.*, 38:3468-4377; page 3472 describes the stability order found by Peyret and coworkers.
[166] Lane, A.; PEack, B.; **(1995)** *Eur. J. Biochem.* 230 1073-1087.
[167] Arnold, F. H.; Wolk, S.; Cruz, P.; Tinoco, I., Jr.; **(1987)** *Biochem.*, 26: 4068-4075.
[168] Oshima, K.; Wells, R.D.; **(1997)** *J. Biol. Chem.*, 272:16798-16806.